MATRIX ALGEBRA for Electrical Engineers BRAAE

MATRIX ALGEBRA

for electrical engineers

R. BRAAE

M.Sc.(Eng.), Hons.B.Sc., M.S.A.I.E.E.

Department of Mathematics, Rhodes University, South Africa

PITMAN

THIS textbook has been written for electrical engineers and will enable them to work in easy stages from first principles to a standard where they will be able to read and understand articles employing matrix techniques. Matrices are appearing with increased frequency in electrical journals and textbooks, and many electrical engineers are realizing that a knowledge of matrix algebra is becoming a pre-requisite if they are to keep abreast of developments in their subject.

The reader will be enabled in this book to visualise the algebraic entities and operations that are introduced and to make the step from matrix algebra to tensor analysis in a simpler and more straight-forward way. For this reason the concepts *transformation, invariance* and *group* are introduced at an early stage and a final chapter is devoted to a brief exposition of the elements of the tensor calculus. The book has been designed for self study and each chapter concludes with a set of problems.

T.1046

30/-
Net

MATRIX ALGEBRA
FOR ELECTRICAL ENGINEERS

MATRIX ALGEBRA

FOR ELECTRICAL ENGINEERS

BY

R. BRAAE

M.Sc. (Eng.), Hons. B.Sc., M.S.A.I.E.E., M.I.E.E.,
Department of Mathematics, Rhodes University,
South Africa

LONDON
SIR ISAAC PITMAN & SONS LTD.

First published 1963

SIR ISAAC PITMAN & SONS LTD.
PITMAN HOUSE, PARKER STREET, KINGSWAY, LONDON, W.C.2
THE PITMAN PRESS, BATH
PITMAN HOUSE, BOUVERIE STREET, CARLTON, MELBOURNE
22–25 BECKETT'S BUILDINGS, PRESIDENT STREET, JOHANNESBURG

ASSOCIATED COMPANIES
PITMAN MEDICAL PUBLISHING COMPANY LTD.
46 CHARLOTTE STREET, LONDON, W.1
PITMAN PUBLISHING CORPORATION
20 EAST 46TH STREET, NEW YORK 17, NEW YORK
SIR ISAAC PITMAN & SONS (CANADA) LTD.
(INCORPORATING THE COMMERCIAL TEXT BOOK COMPANY)
PITMAN HOUSE, 381–383 CHURCH STREET, TORONTO

MADE IN GREAT BRITAIN AT THE PITMAN PRESS, BATH
F3—(T.1046)

To

ANNA GRETE

Preface

I WAS first introduced to matrices as a first-year student of electrical engineering at the University of Copenhagen, and I can still vividly recall my elation when the power and beauty of this fascinating branch of modern algebra dawned upon me. Close and effective liaison must have existed between the departments of pure science, for the algebra taught us by the professors of pure mathematics was immediately employed by their colleagues in the fields of applied mathematics, geometry and physics. Consequently, we were made to realize at a very early stage that matrix algebra is more than an intricate system of abstract symbols. In retrospect I can see, what I did not clearly appreciate at the time, that in many instances we actually wandered into the realm of tensor analysis. More's the pity that the effective analytical tool with which we had been issued was virtually ignored in our subsequent engineering courses.

If I were to emulate the leaders of the French Revolution by coining a slogan such as the famous Liberté : Egalité : Fraternité, in order to epitomize the glories of matrices and tensors, I might employ as my *mots justes* the words Organization : Unification : Visualization.

By means of matrices, discrete elements are combined to form higher entities which are manipulated as units and denoted by single symbols. One advantage of such an organized analysis is the emergence of general principles and concepts which were previously hidden in a jungle of haphazard *ad hoc* symbolisms.

Also, matrices enable many apparently unrelated analytical and physical systems to be brought together under one unified point of view, thus facilitating co-operation between the various branches of pure and applied science. As a result, weapons of analysis and synthesis which have proved their worth in one field can more easily be switched to other fronts. This is all the more important today, seeing that the field of scientific knowledge is expanding so rapidly that it is virtually impossible for anyone, however gifted and diligent, to keep abreast of current advances.

But probably the most useful property of matrices and tensors lies in the fact that operations and concepts that have been defined by purely algebraic means can be visualized in space. For one thing, this is of great mnemonic value. I doubt whether anybody could possibly remember all the theorems given in this book if they had

to be understood simply as algebraic formulae. Having once grasped the geometric significance of a theorem, however, the reader can picture spatially what is being done and there remains very little to remember even in cases where the number of dimensions exceeds that of intuitive three-dimensional space.

Another advantage is the suggestiveness of the geometric approach. Having translated a problem into the language of geometry, we are often able to visualize certain of its analytical difficulties and limitations, or to see new methods of attack which would not be at all obvious to an investigator employing purely algebraic reasoning.

And lastly, geometry provides a universal language by means of which ideas can be exchanged between scientists who would not otherwise understand each other's technical terminology.

In 1953 I was asked to give a series of ten out-of-hours lectures on matrix algebra and its applications at the University of the Witwatersrand in Johannesburg. The very gratifying attendance during the whole of the course seemed to point to a strong latent desire to know more about the subject. This book has grown out of the notes compiled for the course, and they in turn were influenced by my early training in Copenhagen.

In the first place, therefore, thanks are due to my Danish teachers, Professors Jakob Nielsen, A. F. Andersen, Richard Petersen and Børge Jessen. I have often had occasion to be grateful for their thorough and systematic instruction. Secondly, no author describing the electrical applications of matrices and tensors could possibly ignore the inspiration received from Gabriel Kron of Schenectady, U.S.A. Without his pioneering work some of the chapters of this book could not have been written.

Professor A. Heydorn encouraged the work from the very beginning, Professor C. Jacobsz kindly read the manuscript and offered much constructive criticism, Mr. C. H. Badenhorst gave unstintingly of his time, and Mr. P. Jumat performed wonders on the duplicating machine. I have great pleasure in acknowledging this help.

And last but not least I should like to express my gratitude to the publishers for their unfailing helpfulness and courtesy.

It has been my aim in writing this book to let the reader share my enthusiasm for the beauties of matrix algebra and to help him to see the "shape" of the subject. I hope I have succeeded.

R. BRAAE

University of Stellenbosch,
Stellenbosch,
November, 1961

Contents

CHAP.		PAGE
	Preface	vii
1.	INTRODUCTORY	1
	1.1. Scope of the Book	1
	1.2. Nomenclature	2
	1.3. Notation	2
	1.4. Historical Note	3
2.	SCALARS AND VECTORS	6
	2.1. Algebraic Operations with Scalars	6
	2.2. Algebraic Operations with Vectors	7
	2.3. Linear Combinations of Vectors	8
	2.4. Linear Independence of Vectors	8
	2.5. Linear Vector Manifold—Rank	11
	2.6. Linear Equations	12
3.	2-MATRICES AND DETERMINANTS	15
	3.1. 2-Matrices and their Rank	15
	3.2. Permutation of the Positive Integers	16
	3.3. Determinants	18
	3.4. Cofactors	20
	3.5. Minors—Subdeterminants	21
	3.6. Determination of Rank	22
	3.7. Algebraic Operations with Matrices	23
	3.8. Multiplication of Determinants	26
	3.9. The Inverse Matrix	28
4.	GENERAL SOLUTION OF SIMULTANEOUS LINEAR EQUATIONS	33
	4.1. Solution of n Equations with n Unknowns	33
	4.2. General Solution of Simultaneous Linear Equations	35
	4.3. Nullspace of a Matrix	39
	4.4. Union and Intersection of Vector Spaces	41
	4.5. Non-centred Vector Spaces	44
	4.6. Right and Left Inverses of Singular Matrices	45
	4.7. The Factorized Inverse	47

CHAP. PAGE
5. ORTHONORMAL MATRICES—COMPOUND MATRICES . . 51
5.1. Outer Multiplication of Vectors 51
5.2. Scalar Multiplication of Vectors 51
5.3. Orthonormal Matrices 53
5.4. Partitioned Matrices 54
5.5. Partitioned Inverses 56

6. LINEAR TRANSFORMATIONS AND LINEAR VECTOR FUNCTIONS 59
6.1. Coordinate Transformations 59
6.2. Linear Vector Functions. 61
6.3. Rank of a Linear Vector Function 63
6.4. Transformation of Functional Matrix 64
6.5. Definition of a Group—Transformation Groups . 65
6.6. Eigenvectors and Eigenvalues 67
6.7. Invariance of Reduction Polynomial 72
6.8. Reduction of Functional Matrix to Diagonal Form . 74
6.9. The Cayley-Hamilton Theorem 77

7. BILINEAR AND QUADRATIC FORMS 82
7.1. Definition of a Form 82
7.2. Transformation of the Form Matrix 83
7.3. Simultaneous Diagonalization of Two Form Matrices 84
7.4. Equivalence Relations and Equivalence Classes. . 87

8. APPLICATION OF MATRICES TO GEOMETRY AND MECHANICS 90
8.1. Central Quadric Surfaces 90
8.2. Relative Motion 93
8.3. Tensor of Inertia 95

9. LINEAR PROGRAMMING 99
9.1. The Problem of Linear Programming 99
9.2. Definitions 100
9.3. Properties of the Solution Cell K 101
9.4. Dantzig's Simplex Method 103
9.5. Solutions at Infinity—Multiple Maxima—Degeneracy 107

10. LINEAR NETWORK ANALYSIS 110
10.1. Preliminaries and Definitions. 110
10.2. Connexion Matrices—Kirchhoff's Laws 114
10.3. Mesh Method of Analysis 116

CHAP. PAGE
10.4. Node-Pair Method of Analysis 118
10.5. Algebraic Diagram of Network Analysis. . . 122
10.6. Orthogonal Analysis (Mesh-Node Analysis) . . 124

11. DIAKOPTICS 135
11.1. Introduction 135
11.2. Division of Network into Subnetworks . . . 138
11.3. Inversion of Node-Pair Matrix Y′ 140

12. TENSOR ANALYSIS. 145
12.1. Introduction—Non-Linear Coordinate Transformations 145
12.2. Co- and Contravariant Tensors 147
12.3. Tensor Algebra 150
12.4. The Christoffel Symbols 151
12.5. Covariant Differentiation 153
12.6. Intrinsic Differentiation 154
12.7 Conclusion and Prospect 155

Bibliography 159

Index 161

CHAPTER 1

Introductory

1.1. Scope of the Book

DURING the past decade, matrices have been appearing with increasing frequency in electrical journals, and many electrical engineers are realizing that a knowledge of the elements of matrix algebra has become a necessary prerequisite if they wish to keep abreast of developments in many branches of electrical engineering.

This book, written by an electrical engineer for electrical engineers, has been compiled to meet this need. It does not profess to be a highly mathematical treatise dealing rigorously with the concepts underlying modern algebra and then erecting a rigid structure upon a well-prepared foundation; in fact, professional mathematicians will consider many of the proofs to be nothing more than loose demonstrations. Rather, the aim of the author has been to provide a text that will enable an engineer of reasonable mathematical ability and with a modest mathematical background to work his way in easy and convincing steps from first principles to a standard where he will be able to read articles employing matrices without too much difficulty.

In the course of the exposition the concepts *transformation*, *invariance* and *group* are defined, and the theory of matrices is developed in such a way as to dovetail with that of tensors. In this manner a key is provided to the fascinating world of applied tensor analysis pioneered by Gabriel Kron.

The reader is encouraged to adopt a broad view of the subject. Matrix algebra is more than a system of symbols; it provides a higher point of view and enables essentially non-spatial concepts and relationships to be visualized geometrically. This feature is being exploited by various individuals and research teams in an endeavour to make geometry the *universal language* of advanced engineering.

In order to foster a broad approach to our subject we shall not confine ourselves to the purely electrical applications of matrices. Strictly speaking, an engineer can make use of matrix algebra if he knows how to multiply two matrices, understands what is meant by *rank* and can compute the inverse of a non-singular matrix. But

on these iron rations he will find himself ill-equipped to do any original work. For this reason Chapters 8 and 9 have been included. They are not necessary for an understanding of the subsequent chapters on network theory, but serve to give the reader a deeper insight into the nature of matrices by applying them to some non-electrical subjects.

The book is designed for self-study. It contains many exercises to which the serious student can turn his hand in order to gain practice in manipulating matrices. It is particularly important that he should learn to think of a matrix as a single unit rather than as a number of separate elements.

A number of asides have been added in the form of notes. They contain additional details or different points of view which could not readily be incorporated in the main body of the text.

During a first perusal, exercises and notes should be by-passed. This will enable the student to get an impression of the scope of the book before attempting to master its contents in detail. At the second reading every exercise should be treated as a problem and all the notes should be carefully studied. Additional problems will be found at the end of most chapters.

1.2 Nomenclature

A formidable stumbling-block to anyone wishing to acquaint himself with matrix and tensor analysis is the hopeless confusion that exists as regards nomenclature. Not only is there a multiplicity of terms for most concepts, but one and the same word means different things to different authors.

There being no absolute authority on the subject, we shall introduce out own system of terminology, which will, however, follow the lead given by men such as Schouten, Gibbs, Kron and le Corbeiller. To make things easier for those readers who might wish to proceed to more advanced books on the subject, alternative names are given in brackets whenever a new term is introduced. Also, for the sake of style, we shall frequently employ two or more words to cover the same thing. For instance, a square matrix whose determinant is non-zero has an inverse and is said to be either *invertible* or *non-singular*. These two terms will be used at random.

1.3. Notation

Two systems of notation are in common use, *direct* notation and *kernel-index* notation.

In direct notation a matrix is denoted by a single symbol set in bold type. We shall adopt the convention that 1-matrices (vectors) will be denoted by bold italics, whereas 2-matrices are set in bold upright type. Because of the way matrix multiplication is defined the order of the factors in a product is important.

According to the kernel-index system a matrix or tensor is indicated by means of a kernel letter together with upper and/or lower indices (the number of which is called the *valence* of the matrix or tensor). Multiplication or, as some authors term it, compounding, of two quantities is shown by repeating an index (a so-called *dummy* index) and employing Einstein's summation convention which dispenses with the summation sign Σ. Thus

$$a_{ij}x^j \stackrel{\mathrm{D}}{=} \sum_{j=1}^{n} a_{ij}x^j$$

where $\stackrel{\mathrm{D}}{=}$ means "equal to by definition."

When kernel-index notation is used the method of compounding is clearly shown by the position of the indices and the order of the factors is therefore of no significance.

Kernel-index notation is superior to direct notation when quantities of valence higher than 2 are considered. Both systems are essentially shorthand notations, and when actual calculations are to be performed the matrices must be written out in full (or punched on cards) as arrays of numbers or symbols.

We shall employ direct notation except in the chapter on tensor analysis, where the more versatile kernel-index notation will be used. This will allow the reader to master the simpler direct notation which will suffice for most purposes and still acquaint him with the kernel-index system, which he will have to know if he wishes to go on to more advanced texts.

1.4. Historical Note

Descartes's introduction in 1637 of analytical geometry brought about a beautiful organization of algebra and geometry. A point in space was indicated by three numbers (its coordinates with respect to three axes of reference), and a number of well-known algebraic identities could now be interpreted geometrically.

Wessel in 1797 and Argand in 1806 showed how complex numbers might be represented by points in a plane and the algebraic operations of addition, subtraction, multiplication and division given geometric significance.

In 1843–44 the almost simultaneous appearance of Grassmann's *Ausdehnungslehre* and Hamilton's *Quaternions* ushered in an algebra of geometric forms from which sprang among other things the discipline of mathematics known as vector analysis. Grassmann's

theory of extensions was conceived on such a wide scale that it contained the germs of matrix and tensor analysis. Probably because his work was so far ahead of its time, Grassmann never received the recognition he so richly deserved.

Determinants were used by Leibniz (1693) for the solution of simultaneous linear equations. This work was forgotten, however, and only revived in 1750 by the Swiss mathematician Cramer, who, in an appendix to his book, *Introduction à l'analyse des lignes courbes algébriques*, gave the basis for the theory of determinants. Jacobi also made important contributions to this branch of mathematics, and his name is still connected with a particular determinant of derivatives (the Jacobian).

By way of linear substitutions, Cayley was led in 1858 to introduce matrices (the name of which was suggested by his friend and fellow-mathematician Sylvester). The typical row-by-column composition of two matrices developed naturally from the manner in which two substitutions are combined.

Tensor analysis, which embraces most of the ideas and techniques of vector and matrix algebra in addition to concepts culled from the fields of non-Euclidean geometry and the theory of groups, was first expounded by the Italian Ricci in 1888. In 1901 Ricci and a pupil of his, Levi-Civita, published a paper on the *absolute differential calculus* and its applications which, however, did not attract much attention at the time (the name *tensor* is of later origin; it was first used when the calculus was applied to the theory of elasticity).

Einstein's use of tensors in developing his famous models of the universe, special relativity in 1905 and general relativity in 1916, gave considerable impetus to their general acceptance, and Heisenberg's *matrix mechanics* in 1925 focused attention on the field of matrix algebra.

Tensor analysis was applied to analytical dynamics by Synge (1926) in a paper entitled "On the geometry of dynamics."

In the field of electrical engineering, Feldtkeller and Strecker studied the properties of four-poles by means of matrices (1929).

In 1932 Kron entered the arena with the mimeographed work *Tensor Analysis of Rotating Machinery*, and since then a steady flow of papers and books dealing with the analysis of electromechanical, thermal, nuclear and even transportation systems by means of matrix and tensor techniques has come from his hand, culminating in 1953 with the announcement of the principle of *diakoptics*, i.e. the analysis of large and complicated systems by means of *tearing*.

During the last ten years two organizations have been formed for the purpose of studying and promoting tensor and matrix algebra as a *universal language* for engineers. One of these organizations is the Research Association of Applied Geometry, the foundations of which were laid in 1951, when a group of Japanese mathematicians, physicists and engineers led by Professor Kondo of the University of Tokyo commenced a research programme entitled "The Unifying Study of Basic Problems in Engineering and Physical Sciences by Means of Geometry." Since its inception the R.A.A.G. has published two massive volumes containing articles on the application of matrices, tensors and topology to a wide variety of engineering problems.

A second organization, the Tensor Club of Great Britain, was founded in 1950 to promote "the application of determinants, matrices, vectors, dyadics and tensors to the engineering and physical sciences."

As to the future there can be no doubt that geometry will become increasingly important as a common medium of expression by means of which research workers in various fields of physics and engineering will be able to discuss problems which may differ in detail but which are basically isomorphic.

The electrical engineer who is interested in the more advanced theoretical aspects of his subject, but who is ignorant of matrix algebra and tensor calculus, may well find himself at a serious disadvantage.

CHAPTER 2

Scalars and Vectors

2.1. Algebraic Operations with Scalars

WHEN developing a manipulative system of symbols (or, in other words, an algebra) there are certain principles and entities the validity and existence of which must be examined before the algebra can be applied to any practical problems. They are—

(i) The commutative principle.
(ii) The associative principle.
(iii) The existence of an identity (or unit) element.
(iv) The existence of the inverse of an element.

If the algebra permits of more than one operation, the validity of the distributive principle must also be investigated.

In this book we are exclusively concerned with the algebraic operations of addition (subtraction) and multiplication (division). Also, by *number* we shall mean either real or complex numbers.

If we consider addition of numbers, both the commutative and associative principles are valid: $a + b = b + a$, and $(a + b) + c = a + (b + c)$. Zero is the identity element which when added to another element leaves it unchanged: $a + 0 = a$. For any element a there exists a unique additive inverse (opposite) element $-a$ such that $a + (-a) = 0$.

Multiplication of numbers is also both commutative and associative. The identity element is unity, and for each non-zero element a there exists a unique inverse (reciprocal) element a^{-1} such that $a \,.\, a^{-1} = 1$.

Combining the two operations is the distributive principle. Multiplication is distributive with respect to addition; thus

$$a(b + c) = (ab) + (ac)$$

whereas addition is not distributive with respect to multiplication.

It may, however, amuse an electrical engineer to recall that in Boolean contactor algebra perfect symmetry exists between the two operations, so that

$$a + (bc) = (a + b)(a + c)$$

Many of the entities to be introduced later do not comply with all the above-mentioned conditions. It is therefore necessary continually to bear in mind which operations are permissible and which are not.

2.2. Algebraic Operations with Vectors

A collection of n numbers taken in a definite order is called an n-dimensional vector. We shall denote vectors either by writing them out in full as a column of numbers or letters held together by brackets or as one letter set in bold italics. Thus

$$\boldsymbol{A} = \begin{Bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_n \end{Bmatrix}$$

The a's are the coordinates (components) of the vector and n is its dimension. A vector is also termed a 1-matrix (monovalent matrix). In kernel-index notation it is written a_i or a^i, where i is a running (free) index with the range 1 to n.

Two vectors of equal dimension are added by adding corresponding coordinates—

$$\boldsymbol{A} + \boldsymbol{B} = \begin{Bmatrix} a_1 \\ \cdot \\ \cdot \\ a_n \end{Bmatrix} + \begin{Bmatrix} b_1 \\ \cdot \\ \cdot \\ b_n \end{Bmatrix} = \begin{Bmatrix} a_1 + b_1 \\ \cdot \\ \cdot \\ a_n + b_n \end{Bmatrix}$$

An important vector is the *nullvector*, all the components of which are zero; it is denoted by $\boldsymbol{0}$.

$$\boldsymbol{U}_1 = \begin{Bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 0 \end{Bmatrix}, \quad \boldsymbol{U}_2 = \begin{Bmatrix} 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 0 \end{Bmatrix}, \ldots, \quad \boldsymbol{U}_n = \begin{Bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{Bmatrix}$$

are the n n-dimensional unit vectors.

A vector is multiplied by a scalar by multiplying each coordinate—

$$k\boldsymbol{A} = k \begin{Bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_n \end{Bmatrix} = \begin{Bmatrix} ka_1 \\ ka_2 \\ \cdot \\ \cdot \\ ka_n \end{Bmatrix}$$

When defined addition of vectors is commutative and associative, $\mathbf{0}$ is the zero element and $-A$ is the additive inverse of A. Multiplication of a vector by a scalar is distributive. Multiplication of one vector by another will be defined later (*see* Sections 5.1 and 5.2, p. 51).

2.3. Linear Combination of Vectors

When a number of equi-dimensional vectors are multiplied by arbitrary constants and added, the sum is said to be a *linear combination* of the vectors—

$$B = k_1A_1 + k_2A_2 + \ldots + k_mA_m$$

If all the constants are non-negative we speak of a positive linear combination.

Exercise 1. Compute the linear combinations $A - B$ and $A + B - 2C$, where

$$A = \begin{Bmatrix} 3 \\ 0 \\ 3 \\ 2 \end{Bmatrix}, \quad B = \begin{Bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{Bmatrix}, \quad C = \begin{Bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{Bmatrix}$$

Exercise 2. Calculate the linear combination

$$B = k_1U_1 + k_2U_2 + \ldots + k_nU_n$$

where $U_1, U_2, \ldots, U_n$ are the n n-dimensional unit vectors.

2.4. Linear Independence of Vectors

A set of vectors $A_1, A_2, \ldots, A_m$ is said to be *linearly independent* if the vector equation

$$x_1A_1 + x_2A_2 + \ldots + x_mA_m = \mathbf{0}$$

is satisfied when and only when all the coefficients x are zero. Otherwise the vectors are *linearly dependent*.

The three 4-dimensional vectors in Exercise 1, Section 2.3, are linearly dependent since a linear combination of them with non-zero coefficients is equal to the 4-dimensional nullvector.

Exercise 1. It is vitally important that the reader should grasp the significance of the concept of linear independence as soon as possible. At first the idea will mean no more than the literal content of its definition; but gradually he will develop a feel for the concept and be able to visualize in space what he is doing algebraically.

In the case of 3-dimensional vectors in ordinary Euclidean 3-space the conditions for linear independence are set out in Table 2.1.

Table 2.1

CONDITIONS FOR LINEAR INDEPENDENCE OF 3-DIMENSIONAL VECTORS

Number of vectors	Condition for linear independence	All linear combinations will lie in
1	vector must be non-zero	1-space (a line)
2	vectors must be non-collinear	2-space (a plane)
3	vectors must be non-coplanar	3-space ("space")
4 or more	4 or more 3-dimensional vectors are always linearly dependent	3-space

Exercise 2. The n n-dimensional unit vectors are linearly independent, since

$$x_1U_1 + x_2U_2 + \ldots + x_nU_n = \begin{Bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{Bmatrix}$$

and this vector can only equal the n-dimensional nullvector when

$$x_1 = x_2 = \ldots = x_n = 0$$

Exercise 3. It is readily proved that a set of vectors containing the nullvector is linearly dependent. Let the set consist of the vectors $\mathbf{O}, A_1, A_2, \ldots, A_m$. The equation

$$x\mathbf{O} + 0A_1 + 0A_2 + \ldots + 0A_m = \mathbf{0}$$

is satisfied for any value of x. The set is therefore dependent.

Exercise 4. Let a set of linearly independent vectors $A_1, \ldots, A_m$ be given. We shall now prove that any subset of this set will also be linearly independent.

If we assume the subset to be linearly dependent, we have

$$k_1A_{a_1} + k_2A_{a_2} + \ldots + k_pA_{a_p} = \mathbf{0}$$

where $p < m$; $a_1, a_2, \ldots, a_p$ is a subset of the numbers 1 to m; and the k's are not all zero.

Such an assumption would, however, lead to the conclusion that the set of A-vectors were linearly dependent, since the equation

$$k_1A_{a_1} + k_2A_{a_2} + \ldots + k_pA_{a_p} + 0A_{a_{p+1}} + \ldots + 0A_{a_m} = \mathbf{0}$$

would also be satisfied, and this is contrary to our hypothesis. Hence, the subset is linearly independent.

Theorem 1. In a set of linearly dependent vectors at least one vector of the set can be expressed as a linear combination of the others.

Proof. Let $A_1, A_2, \ldots, A_m$ be a set of linearly dependent vectors. Because the vectors are dependent there must exist a set of numbers $x_1, \ldots, x_m$ not all of which are zero, such that

$$x_1A_1 + x_2A_2 + \ldots + x_mA_m = 0$$

Assume for the sake of convenience that $x_1 \neq 0$; then

$$A_1 = -(x_2/x_1)A_2 - (x_3/x_1)A_3 \ldots - (x_m/x_1)A_m$$

which proves the theorem.

Conversely, if one vector in a set can be written as a linear combination of the others, the set is linearly dependent.

Exercise 5. A set of n-dimensional vectors

$$A = \begin{Bmatrix} a_1 \\ \cdot \\ \cdot \\ a_{n-2} \\ a_{n-1} \\ a_n \end{Bmatrix}, \quad B = \begin{Bmatrix} b_1 \\ \cdot \\ \cdot \\ b_{n-2} \\ b_{n-1} \\ 0 \end{Bmatrix}, \quad C = \begin{Bmatrix} c_1 \\ \cdot \\ \cdot \\ c_{n-2} \\ 0 \\ 0 \end{Bmatrix}$$

where $a_n, b_{n-1}, c_{n-2}, \ldots$ are non-zero, is linearly independent.

The linear combination

$$x_aA + x_bB + x_cC + \ldots$$

can only equal the nullvector if its nth coordinate x_aa_n vanishes. This requires $x_a = 0$. To make the $(n-1)$th coordinate vanish, x_b must be zero, and by continuing this train of reasoning we conclude that all the coefficients must be zero, which means that the vectors are independent. The proof hinges on the fact that the number of vectors in the set must be $\leqslant n$.

Let $A_1, A_2, \ldots, A_m$ be a set of vectors. From this set we derive another set as follows—

$$B_1 = A_1, B_2 = A_2 + k_2A_1, \ldots, B_m = A_m + k_mA_1$$

where $k_2, k_3, \ldots, k_m$ are arbitrary constants.

We shall now prove that, if the A-vectors are linearly independent, then so are the B-vectors, and vice versa.

Consider a linear combination of the B-vectors—

$$x_1B_1 + x_2B_2 + \ldots + x_mB_m = x_1'A_1 + x_2A_2 + \ldots + x_mA_m$$

where $x_1' = x_1 + k_2x_2 + \ldots + k_mx_m$.

If we assume the A-vectors to be linearly independent, the linear combination given above will only equal 0 if $x_1' = x_2 = \ldots = x_m = 0$; hence $x_1 = 0$ and the assumption that the B-vectors are dependent will immediately lead to a contradiction.

Conversely, if we let the B-vectors be independent, we can conclude that the A-vectors, too, are independent.

We are now in a position to prove an important theorem.

Theorem 2. A set of m n-dimensional vectors is linearly dependent when $m > n$.

Proof. Let $A_1, \ldots, A_m$ be m n-dimensional vectors and let the nth coordinate of A_1 be non-zero.

It is now possible, as demonstrated above, to derive a set of vectors, $B_1, \ldots, B_m$, which are linearly independent when the set of A-vectors are independent, and dependent when the A's are dependent.

By choosing suitable values for the k's we can make the nth coordinate of the vectors $B_2, \ldots, B_m$ vanish. If, during this process, one of the B-vectors is reduced to the nullvector, our proof is complete. Otherwise, we can repeat the operation with a B-vector the $(n-1)$th coordinate of which is non-zero and obtain $m-2$ vectors $C_3, \ldots, C_m$ the nth and $(n-1)$th coordinates of which are zero.

It is clear that by continuing this process it is always possible to reduce the set of A-vectors to a set containing no more than n vectors of the type discussed in Exercise 5, and the remaining ones to nullvectors. When $m > n$ the derived set will thus contain at least $m - n$ nullvectors and therefore be linearly dependent. Hence, the original set is dependent. This proves the theorem.

Exercise 6. The four 3-dimensional vectors

$$\begin{Bmatrix} 2 \\ 1 \\ 1 \end{Bmatrix}, \quad \begin{Bmatrix} 0 \\ 3 \\ 2 \end{Bmatrix}, \quad \begin{Bmatrix} 3 \\ -2 \\ 1 \end{Bmatrix}, \quad \begin{Bmatrix} 3 \\ 2 \\ 2 \end{Bmatrix}$$

are reduced in the manner just described to

$$\begin{Bmatrix} 2 \\ 1 \\ 1 \end{Bmatrix}, \quad \begin{Bmatrix} -4 \\ 1 \\ 0 \end{Bmatrix}, \quad \begin{Bmatrix} -11 \\ 0 \\ 0 \end{Bmatrix}, \quad \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix}$$

In order to visualize the algebraic operations of this exercise the vectors should be plotted in a perspective sketch.

Note. The act of multiplying a vector by a constant and adding it to another vector is termed a *linear operation*. We can thus look upon the set of B-vectors as being derived from the A-set by means of linear operations. It was proved that two sets of vectors related in this way are either both linearly independent or both linearly dependent.

Paraphrasing this result we can say that the linear independence of a set of vectors cannot be destroyed, or in fact altered in any way, by linear operations.

2.5. Linear Vector Manifold—Rank

The maximum number of linearly independent vectors that can be selected from a finite or infinite set of n-dimensional vectors is called the *rank* of the set. Such a linearly independent subset is said to form a *base* of the set. As we see from Theorem 2, the rank

of a set of vectors can never exceed the dimension of the vectors of the set.

Suppose $V_1, V_2, \ldots, V_r$ to be a base of the set. Then the vectors $V_1, \ldots, V_r, A$, where A is any vector of the set, must be linearly dependent, and A can therefore be written as a linear combination of the base vectors—

$$A = a_1V_1 + \ldots + a_rV_r$$

The a's are the coordinates of A in the frame of reference formed by the base vectors.

The resolution of A into components along the axes of the base is unique. To prove this let us assume that A could be resolved in another way as

$$A = b_1V_1 + b_2V_2 + \ldots + b_rV_r$$

Since $A - A = 0$, we get

$$(a_1 - b_1)V_1 + (a_2 - b_2)V_2 + \ldots + (a_r - b_r)V_r = 0$$

and because the V's are linearly independent, the coefficients of the V's must all vanish and we have $a_1 = b_1$, $a_2 = b_2$, $\ldots$, $a_r = b_r$, which proves the uniqueness of the resolution.

An infinite set of vectors of rank p is called a *linear vector manifold* (a centred vector space, an E_p, a p-direction) if it meets the following conditions—

(i) If it contains the vector A it will also contain the vector kA, where k is any real number.

(ii) If it contains the vectors A and B it will also contain the vector $A + B$.

In short, if the set contains certain vectors, it will contain all linear combinations of them.

The rank r of a vector space is the maximum number of linearly independent vectors that can be selected from it. Such a set forms a base of the manifold, and any vector belonging to the manifold can be uniquely expressed as a linear combination of the base vectors. The manifold is said to be *spanned* (generated) by the base vectors.

Table 2.1 shows a number of manifolds in 3-dimensional space.

2.6. Linear Equations

By employing the concepts just introduced it becomes possible to treat systems of simultaneous linear equations in an organized and systematic manner.

A set of n linear equations with m unknowns is usually written

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m &= b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m &= b_n \end{aligned}$$

The problem to be solved can be stated: Does a set of m numbers $x_1, \ldots, x_m$ exist such that the above n equations will be simultaneously satisfied?

In vector notation the equations can be written much more concisely as

$$x_1A_1 + x_2A_2 + \ldots + x_mA_m = B$$

and the problem can be reworded: Can the vector B be expressed as a linear combination of the vectors $A_1, A_2, \ldots, A_m$?

Apparently the necessary and sufficient condition for a solution is that B shall belong to the vector space spanned by the A-vectors. The great advantage of this approach lies in the fact that the problem can be visualized geometrically. Even when $n > 3$ and we can no longer rely on our normal spatial intuition, geometry is still a very helpful guide.

A general discussion of simultaneous linear equations is given in Sections 4.1 and 4.2, pp. 33 and 35.

PROBLEMS

2.1. Prove that the three 4-dimensional unit vectors U_1, U_2 and U_3 are linearly independent.

2.2. According to Theorem 2, the three 2-dimensional vectors

$$A = \begin{Bmatrix} 1 \\ 2 \end{Bmatrix}, \quad B = \begin{Bmatrix} -2 \\ 0 \end{Bmatrix} \quad \text{and} \quad C = \begin{Bmatrix} 3 \\ 2 \end{Bmatrix}$$

must be linearly dependent. Express each vector as a linear combination of the others and plot the results in a plane coordinate system (not necessarily an orthogonal system).

2.3. Let

$$A_1 = \begin{Bmatrix} 2 \\ 1 \\ 3 \end{Bmatrix}, \quad A_2 = \begin{Bmatrix} -1 \\ 2 \\ 2 \end{Bmatrix} \quad \text{and} \quad A_3 = \begin{Bmatrix} 1 \\ -1 \\ 1 \end{Bmatrix}$$

be 3 vectors in a three-dimensional Cartesian coordinate system. Are these vectors independent? Form a new set of vectors $B_1 = A_1$, $B_2 = A_2 + k_2A_1$, $B_3 = A_3 + k_3A_1$. Sketch both sets and discuss the fact that both sets are either linearly independent or linearly dependent.

2.4. Repeat Problem 2.3 with the following three vectors—

$$A_1 = \begin{Bmatrix} 1 \\ 2 \\ 2 \end{Bmatrix}, \quad A_2 = \begin{Bmatrix} 0 \\ 2 \\ 3 \end{Bmatrix}, \quad \text{and} \quad A_3 = \begin{Bmatrix} 3 \\ 2 \\ 0 \end{Bmatrix}$$

2.5. Let four 4-dimensional vectors

$$A_1 = \begin{Bmatrix} -2 \\ 1 \\ 0 \\ 2 \end{Bmatrix}, \quad A_2 = \begin{Bmatrix} 4 \\ 2 \\ 1 \\ 1 \end{Bmatrix}, \quad A_3 = \begin{Bmatrix} 2 \\ -3 \\ 2 \\ -1 \end{Bmatrix}, \quad \text{and} \quad A_4 = \begin{Bmatrix} 2 \\ -7 \\ 1 \\ x \end{Bmatrix}$$

be given.

Determine x such that the four vectors are linearly dependent. Interpret the result geometrically.

CHAPTER 3

2-Matrices and Determinants

3.1. 2-Matrices and Their Rank

An array of elements consisting of n rows and m columns is called an n-by-m matrix (2-matrix, matrix of valence 2, divalent matrix) and denoted by a_{ij} ($i = 1, \ldots, n$; $j = 1, \ldots, m$) in kernel-index notation, or simply by **A** in direct notation.

Another way of looking at a matrix is to consider it as an aggregate of m n-dimensional column vectors or n m-dimensional row vectors. These two sets of vectors are naturally closely related, consisting as they do of the warp and woof of one and the same array of numbers. This interrelation is given in the next theorem.

Theorem 3. The set of vectors formed by the columns and rows of a matrix are of equal rank. This rank is that of the matrix.

Let the rank of the column vectors be s. Hence s of these vectors will span the set, and by interchanging columns we can move these s vectors so that they occupy the first s columns. This process will permute a number of the coordinates of the row vectors but will in no way affect the rank of the set. Likewise, the r vectors forming a base for the set of row vectors are moved so as to occupy the first r rows. Our problem is now to prove that $s = r$.

We know that the s column vectors are linearly independent, and we shall now prove that, if their dimension is reduced by deleting their last $m - r$ coordinates, they will still be linearly independent.

If we assume that the reduced vectors are linearly dependent, then the equation

$$k_1 \begin{Bmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ a_{r1} \end{Bmatrix} + k_2 \begin{Bmatrix} a_{12} \\ a_{22} \\ \cdot \\ \cdot \\ a_{r2} \end{Bmatrix} + \ldots + k_s \begin{Bmatrix} a_{1s} \\ a_{2s} \\ \cdot \\ \cdot \\ a_{rs} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{Bmatrix} \quad . \tag{3.1}$$

will be satisfied by a non-zero set of k's.

Since the first r row vectors span the complete set, all those in rectangle b of the matrix shown diagrammatically in Fig. 3.1 are linear combinations of the row vectors in a.

As a result the k's of eqn. (3.1) will also satisfy the equation

$$k_1 \begin{Bmatrix} a_{r+1\,1} \\ a_{r+2\,1} \\ \cdot \\ \cdot \\ a_{n\,1} \end{Bmatrix} + k_2 \begin{Bmatrix} a_{r+1\,2} \\ a_{r+2\,2} \\ \cdot \\ \cdot \\ a_{n\,2} \end{Bmatrix} + \ldots + k_s \begin{Bmatrix} a_{r+1\,s} \\ a_{r+2\,s} \\ \cdot \\ \cdot \\ a_{n\,s} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{Bmatrix} \quad (3.2)$$

This in turn means that the s column vectors are linearly dependent, which is contrary to our premiss. Thus the s r-dimensional column vectors are linearly independent, and according to Theorem 2, p. 11, $s \leq r$. By repeating our arguments in terms of the r s-dimensional row vectors we conclude that $r \leq s$. Hence $r = s$, which proves the theorem.

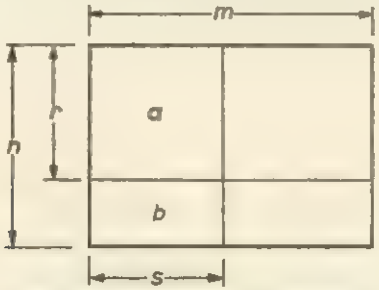


FIG. 3.1. THE RANK OF A MATRIX

This common rank $r\ (= s)$ is that of the matrix. As we have seen, it is invariant to any interchange of rows or columns.

Theorem 4. If the elements in a column (row) of a matrix are augmented by those of another column (row) multiplied by a constant, the rank of the matrix remains unaltered.

After the operation the column (row) vectors will be linear combinations of the columns (rows) of the original matrix and will therefore belong to the same vector space. As we have shown in Section 2.4, p. 10, the new set of vectors will have the same rank as the old set. The rank of the matrix is thus the same as before.

3.2. Permutation of the Positive Integers

Before turning our attention to the theory of determinants there are a few concepts in connexion with the permutation of positive whole numbers which must be introduced.

It is well known that $n!$ permutations can be formed from the positive integers $1, 2, \ldots, n$. Let $p_1, p_2, \ldots, p_r, \ldots, p_s, \ldots p_n$

be one of these permutations. The pair of numbers p_r and p_s is said to be an *inversion* if $p_r > p_s$, i.e. if the two numbers are not written in their natural order. The number of inversions in a given permutation is denoted by $I(p_1, p_2, \ldots, p_n)$, and the permutation is called *odd* or *even* according to whether $I(p_1, p_2, \ldots, p_n)$ is an odd or even number. Thus all permutations fall into two distinct classes.

Exercise. The first three integers can be permuted in $3! = 6$ ways, and the corresponding inversion numbers are given in Table 3.1.

Table 3.1

PERMUTATION OF INTEGERS 1, 2, and 3

Permutation	Inversion Number I	Class
1 2 3	0	even
1 3 2	1	odd
2 1 3	1	odd
2 3 1	2	even
3 1 2	2	even
3 2 1	3	odd

Of these 6 permutations 3 are odd and 3 even.

If p_r and p_{r+1} are two adjacent numbers in a permutation, then $I(p_1, \ldots, p_n)$ will increase or decrease by one when p_r and p_{r+1} are interchanged. This is readily seen, since the inversions formed by p_r and p_{r+1} with the rest of the numbers in the permutation will not be affected by the change. Therefore, when two adjacent numbers in a permutation are interchanged, the permutation changes class.

Let us now consider what the effect will be if two arbitrary numbers, p_r and p_{r+s}, are interchanged. We shall achieve our object in two steps. First p_r changes place with p_{r+1}, then with p_{r+2} and so on, until after s moves it occupies the position formerly taken up by p_{r+s}. Secondly, p_{r+s} is moved back to p_r's original position; this requires $s - 1$ moves. The two numbers have therefore changed place by the interchange of adjacent numbers $2s - 1$ times. Whatever the value of s, this number is odd. So we find that, when two arbitrary numbers in a permutation are interchanged, the permutation changes class.

From this result we can infer that the number of odd and even permutations of the numbers $1, 2, \ldots, n$ must be the same. If the numbers 1 and 2 (say) are interchanged in each and every permutation, we arrive at exactly the same permutations as before simply

taken in a new order, and in doing so every permutation has changed class.

Consider n doublets $(1, p_1)$, $(2, p_2)$, . . ., (n, p_n), where the first digit indicates the number of the doublet, and $p_1, \ldots, p_n$ is some permutation of the integers 1 to n. By interchanging these doublets two at a time it is possible to arrange them according to the second digit as follows—

$$(q_1, 1), (q_2, 2), \ldots, (q_n, n).$$

Thus a certain number of moves has changed $p_1, \ldots, p_n$ into $1, \ldots, n$, and simultaneously $1, \ldots, n$ into $q_1, \ldots, q_n$. From this fact it is easily concluded that the permutations p and q must belong to the same class.

3.3. Determinants

Any square matrix has associated with it a scalar quantity called the *determinant* of the matrix. It is denoted by det (**A**) or $|\mathbf{A}|$. The determinant of an n-by-n matrix is said to be of the nth *order*.

We now define det (**A**) to be the sum of the $n!$ products of n factors chosen in such a way that each term comprises one element from each row and one from each column. The sign of a term is defined to be -1 to the power $I(p_1, \ldots, p_n)$, where $p_1, \ldots, p_n$ are the column indices after the factors have been ordered according to the row indices. Thus

$$\det(\mathbf{A}) \stackrel{D}{=} \Sigma(-1)^{I(p_1, \ldots, p_n)} \cdot a_{1p_1}a_{2p_2} \ldots a_{np_n} \qquad (3.3)$$

To most readers this definition will probably seem highly artificial and arbitrary at first sight. As the theory of matrices is developed, however, the usefulness of determinants will become clear.

Exercise. A 1-by-1 matrix $A = \{a\}$ has a determinant $\det(A) = a$. A 2-by-2 matrix contains $2^2 = 4$ elements—

$$A = \begin{Bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Bmatrix}$$

and its determinant comprises $2! = 2$ terms when expanded—

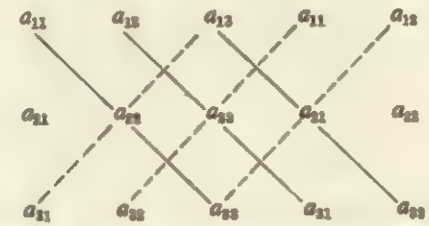
$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

The second term in the expansion is negative because the permutation of the column indices is odd.

A 3-by-3 matrix contains $3^2 = 9$ elements. Its determinant expands into $3! = 6$ terms—

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31}$$
$$+ a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}$$

This particular expansion is easily memorized by the following method. The first and second columns are repeated after the third column as indicated below, and the products along the full (dotted) lines are taken with positive (negative) sign.



It must be borne in mind that this method is *not* valid for determinants of higher order.

Theorem 5. A square matrix and its transpose (the matrix formed by interchanging rows and columns) have identical determinants.

By means of the definition we can expand the determinant of the transpose $\mathbf{A}_t$ of matrix as follows—

$$\det(\mathbf{A}_t) = \Sigma(-1)^{I(q_1, \ldots, q_n)} \cdot a_{q_1 1}a_{q_2 2} \ldots a_{q_n n} \qquad (3.4)$$

It is readily seen that the terms on the right-hand side of eqn. (3.4) are the same as those of eqn. (3.3), except that the latter are ordered according to the last index. Furthermore, since the permutations $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ belong to the same class, the two expansions are identical.

Theorem 6. When all the elements in a row (column) of a determinant are multiplied by a constant k, the determinant is multiplied by k.

This follows from the fact that each term in the expansion contains one factor from each row (and one from each column), so that k can be factorized. A corollary to this theorem is that any determinant, a row (column) of which is zero, also is zero.

Theorem 7. When two rows (columns) of a determinant are interchanged the determinant changes sign.

The numerical values of the terms in the expanded form of the determinant are not affected by the change, and since the permutation $p_1, p_2, \ldots, p_n$ changes class when two numbers are interchanged, it is apparent that each term in the expansion will change its sign.

Theorem 8. It follows directly from Theorem 7 that a determinant in which two columns (rows) are equal (or even proportional) is equal to zero.

3.4. Cofactors

In the expanded form of a determinant det (A), $(n-1)!$ of a total of $n!$ terms will contain the element a_{rs}. The sum of these terms can be written $a_{rs}A_{rs}$, where A_{rs} is called the *cofactor* of the element a_{rs}. Since each term contains only one element from each row and one from each column, A_{rs} is independent of the elements in row r and column s.

Hence the expanded form of a determinant can be condensed as follows—

$$\det(\mathbf{A}) = a_{r1}A_{r1} + a_{r2}A_{r2} + \ldots + a_{rn}A_{rn} \qquad (3.5)$$

If we substitute in **A** the elements in row r by those from another row (say p), we know from Theorem 8 that the determinant of this new matrix vanishes. Thus we immediately get the important identity

$$a_{p1}A_{r1} + a_{p2}A_{r2} + \ldots + a_{pn}A_{rn} = \begin{cases} \det(\mathbf{A}) & \text{when } p = r \\ 0 & \text{when } p \neq r \end{cases}$$

Expressed in words we have—

Theorem 9. The sum of the elements of a row (column) of a determinant each multiplied by its cofactor is equal to the determinant (this is called expanding the determinant about the elements of a row (column)). The sum of the elements of a row (column) of a determinant each multiplied by the cofactor of the corresponding element of another row (column) is equal to zero.

Exercise. The determinant

$$\begin{vmatrix} 2 & 3 & 4 \\ 6 & 7 & 9 \\ 5 & 10 & 8 \end{vmatrix} = 2.7.8 + 3.9.5 + 4.6.10 - 4.7.5 - 2.9.10 - 3.6.8 = 23$$

The cofactor of element $a_{11} = 2$ is $A_{11} = 7.8 - 9.10 = -34$. The other cofactors are easily computed—

$$\begin{Bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{Bmatrix} = \begin{Bmatrix} -34 & -3 & 25 \\ 16 & -4 & -5 \\ -1 & 6 & -4 \end{Bmatrix}$$

and the accuracy of Theorem 9 is easily tested.

With the help of this theorem we are now able to establish a property of determinants which is very useful when calculating the value of a determinant.

Theorem 10. If the elements in a row (column) of a determinant are augmented by the corresponding elements of another row (column) multiplied by a constant, the value of the determinant remains unaltered.

Suppose we multiply the rth row by k and add it to row p. The modified determinant det (A′) expands as follows about row p—

$$\begin{aligned} \det(\mathbf{A}') &= (a_{p1} + ka_{r1})A_{p1} + \ldots + (a_{pn} + ka_{rn})A_{pn} \\ &= (a_{p1}A_{p1} + \ldots + a_{pn}A_{pn}) + k(a_{r1}A_{p1} + \ldots + a_{rn}A_{pn}) \\ &= \det(\mathbf{A}) + k.0 = \det(\mathbf{A}) \qquad \text{(Q.E.D.)} \end{aligned}$$

3.5. Minors—Subdeterminants

From an nth order determinant we can form n^2 determinants of order $n-1$ by deleting one row and one column. These determinants are called *minors*. There is a close relationship between the minor M_{rs} derived by deleting row r and column s and the cofactor A_{rs}.

Theorem 11. The cofactor A_{rs} of an element a_{rs} is equal to $(-1)^{r+s}$ times the minor M_{rs}.

A proof of this theorem is best given in two steps—

(i) First we shall prove that the theorem holds when $r = s = n$.

To determine A_{nn} we write down all the terms in the expansion for det (A) containing the factor a_{nn}. They are

$$\Sigma(-1)^{I(p_1,\ldots,p_{n-1},n)} . a_{p_1 1} \ldots a_{n-1\,p_{n-1}} a_{nn}$$

where $p_1, \ldots, p_{n-1}$ can be permuted in $(n-1)!$ ways.

From the above expression it is readily seen that

$$A_{nn} = \Sigma(-1)^{I(p_1,\ldots,p_{n-1})} . a_{1p_1} \ldots a_{n-1\,p_{n-1}} = M_{nn}$$

Thus in this special case the theorem is valid.

(ii) To prove its general validity we move the element a_{rs} to the position occupied by a_{nn}. First row r is moved by $n-r$ interchanges to the position of row n, and then column s is moved by another $n-s$ steps to the position formerly taken up by column n. The desired change has thus been brought about by $2n-(r+s)$ interchanges. Each time two rows or columns change place the determinant changes sign, and the new determinant is therefore equal to the original one times a factor $(-1)^{2n-(r+s)} = (-1)^{r+s}$.

An identical relation holds between A_{rs} (the cofactor of a_{rs} in the determinant det (A)) and A'_{rs} (the cofactor of a_{rs} in det (A′))—

$$A_{rs} = (-1)^{r+s}A'_{rs}$$

By inspection it is seen that $M_{rs} = M'_{rs}$, and since we have shown under (i) that $A'_{rs} = M'_{rs}$, it follows that

$$A_{rs} = (-1)^{r+s} . M_{rs} \qquad \text{(Q.E.D.)}$$

2—(T.1046)

By means of a chequerboard array of signs (Table 3.2) the value of the factor $(-1)^{r+s}$ is easily visualized.

Table 3.2

CHEQUERBOARD ARRAY OF SIGNS

$r \backslash s$	1	2	3	4
1	+	−	+	−
2	−	+	−	+
3	+	−	+	−
4	−	+	−	+

3.6. Determination of Rank

In Section 3.1, p. 15, we introduced the idea of the rank of a matrix. The following two theorems will demonstrate how the determinants associated with a matrix enable its rank to be calculated.

Theorem 12. An n-by-n matrix is of rank n when and only when its determinant is non-zero.

If the n columns of an n-by-n matrix $\mathbf{A}$ are linearly dependent, it is possible by means of linear column operations such as those described in connexion with the proof of Theorem 2, p. 11, to transform $\mathbf{A}$ into a matrix $\mathbf{A}'$ in which at least one column is a nullvector. Therefore $\det(\mathbf{A}') = 0$, and since $\det(\mathbf{A}') = \det(\mathbf{A})$ we have $\det(\mathbf{A}) = 0$.

It now remains to show that when the rank of $\mathbf{A}$ is n (in future we shall write $r(\mathbf{A}) = n$) then $\det(\mathbf{A}) \neq 0$.

When $r(\mathbf{A}) = n$ the column vectors of the matrix are linearly independent. Let

$$\det(\mathbf{A}) = \begin{vmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdot & \cdot & a_{nn} \end{vmatrix}$$

and suppose that $a_{11} \neq 0$. We now multiply column 1 by suitable constants and add it to the other $n - 1$ columns so as to make their first coordinate vanish. This operation does not alter the rank of the matrix (Theorem 4) nor does it alter the value of $\det(\mathbf{A})$ (Theorem 10).

The transformed determinant is

$$\det(\mathbf{A}') = \begin{vmatrix} a_{11} & 0 & \cdot & \cdot & 0 \\ a_{21} & a'_{22} & \cdot & \cdot & a'_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a'_{n2} & \cdot & \cdot & a'_{nn} \end{vmatrix}$$

the n column vectors of which are linearly independent. The $n - 1$ last columns of $\det(\mathbf{A}')$ form a subset which is also linearly independent (see Exercise 4, Section 2.4). Thus the $(n-1)$-by-$(n-1)$ matrix

$$\mathbf{B} = \begin{Bmatrix} a'_{22} & \cdot & \cdot & a'_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a'_{n2} & \cdot & \cdot & a'_{nn} \end{Bmatrix}$$

will be of rank $n - 1$. By expanding $\det(\mathbf{A}')$ about its first row and noting that $\det(\mathbf{B}) = A_{11}$ (the cofactor of a_{11}), we see that

$$\det(\mathbf{A}) = \det(\mathbf{A}') = a_{11} \,.\, \det(\mathbf{B})$$

If we therefore assume that $\det(\mathbf{B}) \neq 0$ then $\det(\mathbf{A})$, too, is non-zero. Since a 1-by-1 matrix is of rank 1 when its determinant is non-zero, the theorem must by induction hold in all cases.

Theorem 13. An n-by-m matrix is of rank r if it contains at least one non-zero subdeterminant of the rth order but no non-zero subdeterminant of higher order.

By linear operations which do not affect rank, the matrix can be transformed into another with r linearly independent columns occupying the first r columns, r independent rows occupying the first r rows, and all the other elements equal to zero. From this second matrix the theorem is immediately apparent by inspection.

3.7. Algebraic Operations with Matrices

In this section we shall develop an algebra of matrices by defining the operations of addition and multiplication. Before doing so, however, we must introduce some new terms.

Two matrices with the same number of rows and columns are said to be *like*. A *nullmatrix* is one in which all the elements are zero. Nullmatrices are denoted by $\mathbf{0}$, and it is left to the reader to conclude from the context the particular shape of the nullmatrix in question.

Like matrices are added by adding corresponding elements. This process is commutative and associative. A zero element ($\mathbf{0}$) and the opposite of a matrix exist.

A matrix is multiplied by a constant k by multipying each of its elements by k. This operation is commutative and associative. Also, If $\mathbf{A}$ is an n-by-n matrix,

$$\det(k\mathbf{A}) = k^n \det(\mathbf{A}) \qquad . \quad . \quad . \ (3.6)$$

A square matrix in which the elements along the *main diagonal* (top left to bottom right) are unity, and the remainder are zero, is called a *unit matrix* (identity matrix, idemfactor, Kronecker delta). It is written

$$\mathbf{I} = \begin{Bmatrix} 1 & 0 & 0 & . & . & 0 \\ 0 & 1 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & 1 \end{Bmatrix}$$

also $\det(\mathbf{I}) = 1$.

A square matrix which has non-zero elements along only the main diagonal is called a *diagonal matrix*. Its determinant is equal to the product of the diagonal elements.

By interchanging the rows and columns of a matrix $\mathbf{A}$ we derive its *transpose*, which is written $\mathbf{A}_t$. When $\mathbf{A}_t = \mathbf{A}$, the matrix is said to be *symmetric* (self-conjugate), and when $\mathbf{A}_t = -\mathbf{A}$ it is called *skew* (skew symmetric, anti-selfconjugate, alternating). Symmetric and skew matrices are necessarily square and the latter have zero elements along the main diagonal.

Multiplication of one matrix by another (also termed the *compounding* of two matrices) can be defined in a number of different ways. We shall proceed as follows.

Two matrices can be multiplied when the number of columns in the *pre-factor* is equal to the number of rows in the *post-factor*. The element in the rth row and sth column of the product is equal to the sum of the elements in the rth row of the pre-factor each multiplied by the corresponding element in the sth column of the post-factor. Expressed algebraically, if $\mathbf{C} = \mathbf{AB}$, then

$$c_{rs} = a_{r1}b_{1s} + a_{r2}b_{2s} + \dots + a_{rn}b_{ns} = a_{ri}b_{is}$$

It can easily be seen that the product matrix will have the same number of rows as the pre-factor, and the same number of columns as the post-factor.

Matrix multiplication is generally non-commutative (hence the necessity for the terms pre- and post-factor), even in cases where both $\mathbf{AB}$ and $\mathbf{BA}$ are meaningful. During the early 1840s, while developing his algebra of quaternions, Hamilton was disturbed to find that he would have to abandon the commutative principle.

It took him a long time to realize that a non-commutative algebra could still be self-consistent.

Exercise 1.

Let $\quad \mathbf{A} = \begin{Bmatrix} 1 & 2 \\ 0 & 2 \end{Bmatrix}$ and $\mathbf{B} = \begin{Bmatrix} -1 & 1 \\ 2 & 1 \end{Bmatrix}$

Then $\quad \mathbf{AB} = \begin{Bmatrix} 1 & 2 \\ 0 & 2 \end{Bmatrix} \begin{Bmatrix} -1 & 1 \\ 2 & 1 \end{Bmatrix} = \begin{Bmatrix} 3 & 3 \\ 4 & 2 \end{Bmatrix}$

and $\quad \mathbf{BA} = \begin{Bmatrix} -1 & 1 \\ 2 & 1 \end{Bmatrix} \begin{Bmatrix} 1 & 2 \\ 0 & 2 \end{Bmatrix} = \begin{Bmatrix} -1 & 0 \\ 2 & 6 \end{Bmatrix} \neq \mathbf{AB}$

Exercise 2. Compute $\mathbf{IA}$ and $\mathbf{AI}$, where

$$\mathbf{A} = \begin{Bmatrix} 1 & 1 & 2 \\ -1 & 0 & 1 \\ 2 & 1 & 2 \end{Bmatrix}$$

$$\mathbf{IA} = \begin{Bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Bmatrix} \begin{Bmatrix} 1 & 1 & 2 \\ -1 & 0 & 1 \\ 2 & 1 & 2 \end{Bmatrix} = \begin{Bmatrix} 1 & 1 & 2 \\ -1 & 0 & 1 \\ 2 & 1 & 2 \end{Bmatrix} = \mathbf{A}$$

and
$$\mathbf{AI} = \begin{Bmatrix} 1 & 1 & 2 \\ -1 & 0 & 1 \\ 2 & 1 & 2 \end{Bmatrix} \begin{Bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Bmatrix} = \begin{Bmatrix} 1 & 1 & 2 \\ -1 & 0 & 1 \\ 2 & 1 & 2 \end{Bmatrix} = \mathbf{A}$$

This exercise illustrates two points: first that the idemfactor actually is a *unit* matrix, and secondly that matrix multiplication may be commutative in special cases.

Exercise 3. Compute $\mathbf{AB}$ and $\mathbf{BA}$, where

$$\mathbf{A} = \begin{Bmatrix} 1 & 2 \\ 2 & 4 \end{Bmatrix} \text{ and } \mathbf{B} = \begin{Bmatrix} 2 & -6 \\ -1 & 3 \end{Bmatrix}$$

$$\mathbf{AB} = \begin{Bmatrix} 1 & 2 \\ 2 & 4 \end{Bmatrix} \begin{Bmatrix} 2 & -6 \\ -1 & 3 \end{Bmatrix} = \begin{Bmatrix} 0 & 0 \\ 0 & 0 \end{Bmatrix} = \mathbf{0}$$

and
$$\mathbf{BA} = \begin{Bmatrix} 2 & -6 \\ -1 & 3 \end{Bmatrix} \begin{Bmatrix} 1 & 2 \\ 2 & 4 \end{Bmatrix} = \begin{Bmatrix} -10 & -20 \\ 5 & 10 \end{Bmatrix} \neq \mathbf{0}$$

This exercise demonstrates that we cannot infer from $\mathbf{AB} = \mathbf{0}$ and $\mathbf{A} \neq \mathbf{0}$ that $\mathbf{B} = \mathbf{0}$.

Exercise 4. Prove that two like diagonal matrices commute.

Let $\mathbf{F}$ and $\mathbf{G}$ be two n-by-n diagonal matrices with non-zero elements f_{ii} and g_{ii} respectively.

The element in row r and column s of the product matrix $\mathbf{FG}$ is zero when $r \neq s$, and equal to $f_{rr}g_{rr}$ (no summation) when $r = s$. The same applies to the elements of $\mathbf{GF}$, and therefore $\mathbf{FG} = \mathbf{GF}$.

We have seen that matrix multiplication is generally non-commutative. Matrix products are, however, always associative.

Suppose we have

$$(\mathbf{A}_{mn}\mathbf{B}_{np})\mathbf{C}_{pq} = \mathbf{D}_{mq}$$

and
$$\mathbf{A}_{mn}(\mathbf{B}_{np}\mathbf{C}_{pq}) = \mathbf{D}'_{mq}$$

where the subscripts indicate the number of rows and columns (respectively) of the matrices.

In order to prove that $\mathbf{D} = \mathbf{D}'$ we shall compute the general term of the two matrices.

The rth row of **AB** contains the elements

$$a_{ri}b_{i1}, \quad a_{ri}b_{i2}, \quad \ldots, \quad a_{ri}b_{ip}$$

Hence
$$d_{rs} = a_{ri}b_{ij}c_{js}$$

where we employ Einstein's summation convention, and the free indices i and j run from 1 to n and from 1 to p respectively.

When calculating d'_{rs} we first compute the elements in the sth column of **BC**; they are

$$b_{1j}c_{js}, \ b_{2j}c_{js}, \ \ldots, \ b_{nj}c_{js}$$

and from this we get d'_{rs} by multiplication by the elements in row r of **A**. Hence

$$d'_{rs} = a_{ri}b_{ij}c_{js} = d_{rs}$$

which proves the associativity of the matrix product. It is thus possible to write $\mathbf{D} = \mathbf{ABC}$ without fear of ambiguity.

The distributive principle holds good for matrix multiplication—

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

and
$$(\mathbf{D} + \mathbf{E})\mathbf{F} = \mathbf{DF} + \mathbf{EF}$$

It should be carefully noted, however, that the order of the factors must not be disturbed.

A useful theorem which is very easy to derive (the proof is left to the reader) is given below.

Theorem 14. The transpose of a matrix product is equal to the product of the transposed factors taken in reverse order. Thus

$$(\mathbf{AB})_t = \mathbf{B}_t\mathbf{A}_t$$

The conditions under which the inverse of a matrix exists will be fully discussed in Section 3.9.

3.8. Multiplication of Determinants

Let **A** and **B** be two n-by-n matrices. Their product **AB** will also be an n-by-n matrix. We shall now prove the following theorem.

Theorem 15. The determinant of the product of two square matrices is equal to the product of the determinants of the factors. Thus

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$$

If two rows in **A** are interchanged the corresponding rows in **AB** are similarly transposed. Also, if a row in **A** is multipled by a constant and added to another row, exactly the same operation will be performed with the corresponding rows in **AB**. These linear operations leave both det (**A**) and det (**AB**) unchanged.

We subdivide the proof into two cases—

(i) det (**A**) = 0. In this case the n rows of **A** are linearly dependent and at least one of them is a linear combination of the others. By linear operations we can reduce this row vector to $\boldsymbol{0}$, and since corresponding row in **AB** will also consist of zeros, det (**AB**) = 0 and the theorem has been proved under these particular conditions.

(ii) det (**A**) $\neq 0$. The rows of **A** are now independent, and by linear row-operations alone we can reduce **A** to diagonal form. This process affects neither det (**A**) nor det (**AB**).

Denoting the diagonalized matrix by **A**′, we have

$$\mathbf{A}' = \begin{Bmatrix} a'_{11} & 0 & . & . & 0 \\ 0 & a'_{22} & & & \\ . & . & & & \\ . & . & . & . & . \\ 0 & 0 & . & . & a'_{nn} \end{Bmatrix}$$

and

$$\mathbf{B} = \begin{Bmatrix} b_{11} & b_{12} & . & . & b_{1n} \\ b_{21} & b_{22} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ b_{n1} & b_{n2} & . & . & b_{nn} \end{Bmatrix}$$

so that

$$\det(\mathbf{A}'\mathbf{B}) = \begin{vmatrix} a'_{11}b_{11} & a'_{11}b_{12} & . & . & a'_{11}b_{1n} \\ a'_{22}b_{21} & a'_{22}b_{22} & . & . & a'_{22}b_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ a'_{nn}b_{n1} & a'_{nn}b_{n2} & . & . & a'_{nn}b_{nn} \end{vmatrix}$$

In this determinant the a'_{ii}'s can be factorized (Theorem 6, p. 19) and this leads to

$$\begin{aligned} \det(\mathbf{A}'\mathbf{B}) = \det(\mathbf{AB}) &= a'_{11} \,.\, a'_{22} \ldots a'_{nn} \,.\, \det(\mathbf{B}) \\ &= \det(\mathbf{A}) \det(\mathbf{B}) \end{aligned}$$

which proves the theorem.

Exercise. Let **A** and **B** be two n-by-n matrices and let det (**B**) $\neq$ **0**. Prove that **A** and **AB** are of equal rank.

We loose nothing in the way of generality by assuming **A** to be a diagonal matrix in which $r (\leqslant n)$ of the diagonal elements are non-zero. The row vectors

of **AB** will consist of r of the row vectors of **B** (which are linearly independent) multiplied by the non-zero diagonal elements of **A**. This set of vectors is of rank r and the rank of **AB** is also r.

3.9. The Inverse Matrix

In scalar algebra the reciprocal (inverse) of a number was introduced by posing the question, does a number x exist such that $x \,.\, a = 1$? The answer was that x existed provided that $a \neq 0$. In matrix algebra the problem is complicated by the much greater variety of "numbers" and by the fact that matrix multiplication is non-commutative.

We shall begin, therefore, by seeking the *right inverse* of a square matrix **A**, i.e. the matrix **X** which satisfies the equation

$$\mathbf{AX} = \mathbf{I} \qquad \ldots \qquad (3.7)$$

By Theorem 15 we conclude that

$$\det(\mathbf{A}) \,.\, \det(\mathbf{X}) = \det(\mathbf{I}) = 1$$

or

$$\det(\mathbf{X}) = 1/\det(\mathbf{A})$$

so that **X** only exists if $\det(\mathbf{A}) \neq 0$. In Section 3.4, p. 20, it was proved that

$$a_{p1}A_{r1} + a_{p2}A_{r2} + \ldots + a_{pn}A_{rn} = \begin{cases} \det(\mathbf{A}) & \text{when } p = r \\ 0 & \text{when } p \neq r \end{cases}$$

and this means that the matrix

$$\mathbf{X} = 1/\det(\mathbf{A}) \,.\, \begin{Bmatrix} A_{11} & A_{21} & \cdot & \cdot & A_{n1} \\ A_{12} & A_{22} & \cdot & \cdot & A_{n2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ A_{1n} & A_{2n} & \cdot & \cdot & A_{nn} \end{Bmatrix} \qquad \ldots \qquad (3.8)$$

will satisfy eqn. (3.7).

The matrix given in eqn. (3.8) will also be a *left* inverse of **A**. Pre-multiplication of eqn. (3.7) by X yields

$$\mathbf{X(AX)} = \mathbf{(XA)X} = \mathbf{X}$$

which clearly shows that $\mathbf{XA} = \mathbf{I}$.

We can now formulate another theorem.

Theorem 16. When the determinant of a square matrix **A** is non-zero the equations $\mathbf{AX} = \mathbf{I}$ and $\mathbf{XA} = \mathbf{I}$ have a unique solution which is called the inverse (reciprocal) of **A** and is written $\mathbf{A}^{-1}$. The elements of $\mathbf{A}^{-1}$ are given by eqn. (3.8). Furthermore, $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$.

An identity analogous to that of Theorem 14 is given in the next theorem.

Theorem 17. The inverse of the product of two like square matrices is equal to the product of their inverses taken in reverse order. Thus

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

Proof. At this stage we have gained sufficient mastery of the technique of matrix algebra to be able to prove this theorem without having recourse to writing the matrices out in full. Hence

$$(\mathbf{AB})^{-1}(\mathbf{AB}) = \mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{AB} = \mathbf{B}^{-1}\mathbf{IB} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$$

and, as an additional (but unnecessary) check,

$$(\mathbf{AB})(\mathbf{AB})^{-1} = \mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}$$

which proves the theorem.

Exercise 1. A diagonal matrix (say)

$$\mathbf{A} = \begin{Bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{Bmatrix}$$

has an inverse

$$\mathbf{A}^{-1} = \begin{Bmatrix} 1/a & 0 & 0 \\ 0 & 1/b & 0 \\ 0 & 0 & 1/c \end{Bmatrix}$$

as can readily be shown by direct multiplication. The expression is easily generalized to matrices of higher order.

Exercise 2. The inverse of the unit matrix **I** is the unit matrix itself (as, naturally, one would expect it to be).

When calculating the reciprocal of a matrix, the work should be undertaken systematically as follows.

Step 1. Calculate the determinant of the matrix **A**. If $\det(\mathbf{A}) = 0$ no inverse exists.

Step 2. Transpose the matrix, i.e. interchange rows and columns.

Step 3. Replace each element a_{rs} by its minor M_{rs} (M_{rs} is the subdeterminant derived from $\det(\mathbf{A}_t)$ by deleting row r and column s)).

Step 4. Multiply each minor by $+1$ or -1 according to the chequerboard array of signs (see Table 3.2, p. 22). The resulting matrix is related to **A** by being the matrix in which the components are the cofactors of the corresponding components of $\mathbf{A}_t$; this matrix is often called the *adjunct* of **A** ($\mathbf{A}_{adj}$).

Step 5. Finally, each element of the adjunct is multiplied by $1/\det(\mathbf{A})$.

Exercise 3. Calculate the inverse of

$$\mathbf{A} = \left\{\begin{matrix} 2 & 1 & -2 \\ 0 & 1 & 1 \\ -1 & 0 & 0 \end{matrix}\right\}$$

Step 1. det (A) = −3, so that $\mathbf{A}^{-1}$ is defined.

Step 2. Transpose the matrix—

$$\mathbf{A}_t = \left\{\begin{matrix} 2 & 0 & -1 \\ 1 & 1 & 0 \\ -2 & 1 & 0 \end{matrix}\right\}$$

Step 3. Replace each element of $\mathbf{A}_t$ by its minor—

$$\left\{\begin{matrix} 0 & 0 & 3 \\ 1 & -2 & 2 \\ 1 & 1 & 2 \end{matrix}\right\}$$

Step 4. Change the signs of the elements of the matrix of minors according to the chequerboard pattern—

$$\mathbf{A}_{adj} = \left\{\begin{matrix} 0 & 0 & 3 \\ -1 & -2 & -2 \\ 1 & -1 & 2 \end{matrix}\right\}$$

Step 5. Multiply $\mathbf{A}_{adj}$ by 1/det (A) to get $\mathbf{A}^{-1}$—

$$\mathbf{A}^{-1} = \frac{1}{3}\left\{\begin{matrix} 0 & 0 & -3 \\ 1 & 2 & 2 \\ -1 & 1 & -2 \end{matrix}\right\}$$

As a final check, $\mathbf{A}\mathbf{A}^{-1}$ or $\mathbf{A}^{-1}\mathbf{A}$ should be computed.

The following theorem proves that the operations transposition and inversion commute.

Theorem 18. The inverse of the transpose of a (square) matrix is equal to the transpose of the inverse of the matrix. Thus

$$(\mathbf{A}_t)^{-1} = (\mathbf{A}^{-1})_t \quad (= \mathbf{A}_t^{-1})$$

Proof. Post-multiplication of $\mathbf{A}_t$ by $(\mathbf{A}^{-1})_t$ yields $\mathbf{A}_t(\mathbf{A}^{-1})_t = (\mathbf{A}^{-1}\mathbf{A})_t = \mathbf{I}_t = \mathbf{I}$ (by Theorem 14), which proves the theorem. It is left to the reader to convince himself that $(\mathbf{A}^{-1})_t$ is also the left inverse of $\mathbf{A}_t$.

The question of the inverse of a singular matrix, i.e. a non-square matrix or one with zero determinant, will be discussed in Section 4.6, p. 45.

PROBLEMS

3.1. Show by detailed calculation that eqn. (3.2) follows from eqn. (3.1) when we make use of the fact that the first r row vectors of the matrix span the complete set of row vectors. *Hint.* Write the $(r + 1)$th s-dimensional row vector as a linear combination of the r s-dimensional row vectors of rectangle a, Fig. 3.1, and substitute in eqn. (3.2).

3.2 Discuss the statement in Exercise 3, Section 3.7, in the light of Theorem 15.

3.3. Prove that det (−A) = $(-1)^n$ det (A), where is an n-by-n matrix.

3.4. Give a general proof that pre- and post-multiplication of a square matrix by the like unit matrix leaves the matrix unchanged: **AI** = **IA** = **A**.

3.5. Extend Theorems 14 and 17 to any number of matrices.

3.6. Prove that when a matrix is symmetric, then so is its inverse.

3.7. Investigate the conditions under which **A** and **D** commute, where **D** is a diagonal matrix and **A** a non-diagonal, like matrix.

3.8. Let **A** be an n-by-n matrix and let $\mathbf{A}(\mathbf{A} - \mathbf{I}) = \mathbf{0}$. Prove that

$$r(\mathbf{A}) + r(\mathbf{A} - \mathbf{I}) = n.$$

3.9. Let **A** and **B** be square non-singular matrices. Prove that if these commute, then so do $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$, and also **A** and $\mathbf{B}^{-1}$.

3.10. Suppose that the variables y_1 and y_2 are linear functions of the variables x_1 and x_2:

$$y_1 = x_1 - x_2$$
$$y_2 = 2x_1 - x_2$$

Derive by elementary methods a set of equations giving the x_i as functions of the y_i.

3.11. Express the linear substitution given in Problem 3.10 in matrix form, $Y = \mathbf{A}X$, and discuss its invertibility.

3.12. Let the x_i of Problem 3.10 be linear functions of a set of variables z_i:

$$x_1 = 2z_1 + z_2$$
$$x_2 = -z_1 - z_2$$

Solve the equations for z_i by elementary means. Express by direct substitution the y_i as functions of the z_i, where the y_i are the variables mentioned in Problem 3.10.

3.13. Express the linear substitution of Problem 3.12 in matrix form, $X = \mathbf{B}Z$. Determine a matrix **C** such that $Y = \mathbf{C}Z$. Is this substitution invertible? If so, calculate the inverse. Note how the consecutive application of linear substitutions leads naturally to our definition of matrix multiplication (Section 3.7, p. 24).

3.14. Can the following substitution be inverted?

$$y_1 = x_1 + x_2$$
$$y_2 = -x_1 + 2x_2$$
$$y_3 = 2x_1 - x_2$$

3.15. Combine the substitutions given in Problems 3.14 and 3.12, and check that the result tallies with that derived by matrix multiplication.

3.16. Test the validity of the associative principle by calculating (**AB**)**C** and **A**(**BC**), where

(i) $\mathbf{A} = \left\{\begin{matrix} 2 & 0 & -1 \\ 1 & 1 & 2 \\ 2 & -1 & 0 \end{matrix}\right\}, \quad \mathbf{B} = \left\{\begin{matrix} 2 \\ 3 \\ -1 \end{matrix}\right\}, \quad \mathbf{C} = \{1 \quad 0 \quad -1\}$

(ii) $\mathbf{A} = \left\{\begin{matrix} -1 \\ 2 \\ 1 \end{matrix}\right\}, \quad \mathbf{B} = \{1 \quad -3 \quad 6\}, \quad \mathbf{C} = \left\{\begin{matrix} -2 \\ 1 \\ 1 \end{matrix}\right\}$

3.17. Test the validity of the associative principle by computing the values of $V_t((WV_t)W)$ and $(V_tW)(V_tW)$, where

$$V_t = (1 \quad 0 \quad 2 \quad -1) \quad \text{and} \quad W_t = (2 \quad 1 \quad -1 \quad 2)$$

Also, verify that $V_tWV_tW = V_tWW_tV$.

3.18. Show that the product of two scalar products can be rewritten as indicated—

$$(P_tQ)(R_tS) = P_t(R_tSI)Q$$

and test the result by substituting the values $P_t = (-1 \quad 1 \quad 1)$, $Q_t = (2 \quad 1 \quad 2)$, $R_t = (-1 \quad 2 \quad 1)$ and $S_t = (-1 \quad 0 \quad 1)$.

Compare this result with the derivation of eqn. (8.12), p. 96.

3.19. Calculate **AB** and **BA**, and show that

$$\det(\mathbf{AB}) = \det(\mathbf{BA}) = \det(\mathbf{A})\det(\mathbf{B}),$$

where $$\mathbf{A} = \begin{Bmatrix} 1 & 0 & 2 \\ -1 & 2 & 1 \\ 0 & 1 & 1 \end{Bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{Bmatrix} 2 & -1 & 2 \\ 1 & 0 & 3 \\ 1 & 2 & 5 \end{Bmatrix}$$

3.20. Show that $\mathbf{AA}_t$ and $\mathbf{A}_t\mathbf{A}$ are symmetric if **A** is any square matrix.

3.21. Let **A** be a skew n-by-n matrix. Prove that $\mathbf{A}^n$ is symmetric when n is even, and skew (or anti-symmetric) when n is odd.

CHAPTER 4

General Solution of Simultaneous Linear Equations

4.1. Solution of n Equations with n Unknowns

WE have now developed all the mathematical tools required to attack the problem foreshadowed in Section 2.6, p. 13.

As a simple preliminary problem let us take a system of n linear equations with n unknowns, $x_1, \ldots, x_n$.

$$\left.\begin{matrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \cdot \quad \cdot \quad \cdots \quad \cdot \quad \cdot\cdot \\ \cdot \quad \cdot \quad \cdots \quad \cdot \quad \cdot\cdot \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n \end{matrix}\right\} \quad . \quad . \quad (4.1)$$

By considering the coefficients of x_i as a column vector A_i we are able to condense eqns. (4.1) to

$$x_1A_1 + x_2A_2 + \ldots + x_nA_n = B \quad . \quad . \quad (4.2)$$

and to formulate the problem in the language of geometry: Can the n-dimensional vector B be expressed as a linear combination of the n n-dimensional vectors A_i?

If we assume that the n A-vectors are linearly independent, they span n-space, which must contain B, and B can therefore be written as a unique linear combination of the base vectors A_i. The equations have thus one and only one solution.

With the help of matrix multiplication, eqn. (4.2) can be set out even more concisely as

$$\mathbf{A}X = B \quad . \quad . \quad . \quad . \quad (4.3)$$

where **A** is the matrix formed by joining together the n A-vectors to form an n-by-n matrix **A**.

To find the solution we simply pre-mulitply eqn. (4.3) by $\mathbf{A}^{-1}$ (which is defined, since $\det(\mathbf{A}) \neq 0$) and thus get

$$\mathbf{A}^{-1}\mathbf{A}X = \mathbf{I}X = X = \mathbf{A}^{-1}B \quad . \quad . \quad . \quad (4.4)$$

In full, this equation reads

$$\begin{Bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{Bmatrix} = \frac{1}{\det(\mathbf{A})} \begin{Bmatrix} A_{11} & A_{21} & \cdot & \cdot & A_{n1} \\ A_{12} & A_{22} & \cdot & \cdot & A_{n2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ A_{1n} & A_{2n} & \cdot & \cdot & A_{nn} \end{Bmatrix} \begin{Bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_n \end{Bmatrix}$$

$$= \frac{1}{\det(\mathbf{A})} \begin{Bmatrix} b_1A_{11} + b_2A_{21} + \ldots + b_nA_{n1} \\ b_1A_{12} + b_2A_{22} + \ldots + b_nA_{n2} \\ \cdot \quad \cdot \quad \cdots \quad \cdot \\ \cdot \quad \cdot \quad \cdots \quad \cdot \\ b_1A_{1n} + b_2A_{2n} + \ldots + b_nA_{nn} \end{Bmatrix} \tag{4.5}$$

Thus $\qquad x_i = (b_1A_{1i} + b_2A_{2i} + \ldots + b_nA_{ni})/\det(\mathbf{A})$

From Theorem 9, p. 20, it is clear that the numerator of this expression is derived from the determinant of **A** by substituting vector $\boldsymbol{B}$ for column $\boldsymbol{A}_i$. We now recapitulate the above results.

Theorem 19 (*Cramer's Theorem*). When the determinant det (A) belonging to a system of n linear equations with n unknowns is non-zero, the system has a unique solution given by

$$\boldsymbol{X} = \mathbf{A}^{-1}\boldsymbol{B}$$

or $\qquad x_i = \det(\mathbf{B}_i)/\det(\mathbf{A})$

where det ($\mathbf{B}_i$) is the determinant obtained when vector $\boldsymbol{B}$ is substituted for column $\boldsymbol{A}_i$ in det (A).

Exercise. Solve the equations

$$\begin{aligned} x + y + z &= 2 \\ 2x \qquad + 2z &= 0 \\ 2x - 2y + z &= -3 \end{aligned}$$

Written in matrix form they become

$$\begin{Bmatrix} 1 & 1 & 1 \\ 2 & 0 & 2 \\ 2 & -2 & 1 \end{Bmatrix} \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} 2 \\ 0 \\ -3 \end{Bmatrix}$$

det (A) = 2 ($\neq 0$); therefore $\mathbf{A}^{-1}$ exists and the equations have a unique solution. $\mathbf{A}^{-1}$ is calculated as described in Section 3.9, p. 28.

$$\mathbf{A}^{-1} = \tfrac{1}{2} \begin{Bmatrix} 4 & -3 & 2 \\ 2 & -1 & 0 \\ -4 & 4 & -2 \end{Bmatrix}$$

Hence

$$\boldsymbol{X} = \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \tfrac{1}{2} \begin{Bmatrix} 4 & -3 & 2 \\ 2 & -1 & 0 \\ -4 & 4 & -2 \end{Bmatrix} \begin{Bmatrix} 2 \\ 0 \\ -3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 2 \\ -1 \end{Bmatrix}$$

If, however, we are concerned only with (say) the unknown y, we apply the second part of Cramer's theorem and get

$$y = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 2 & 0 & 2 \\ 2 & -3 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ 2 & 0 & 2 \\ 2 & -2 & 1 \end{vmatrix}} = \frac{4}{2} = 2$$

which tallies with the result obtained by matrix multiplication.

4.2. General Solution of Simultaneous Linear Equations

In general, n linear equations with m unknowns can be written vectorially as

$$x_1\boldsymbol{A}_1 + x_2\boldsymbol{A}_2 + \ldots + x_m\boldsymbol{A}_m = \boldsymbol{B} \tag{4.6}$$

where the $\boldsymbol{A}_i$ and $\boldsymbol{B}$ are n-dimensional vectors.

The conditions of solvability are given in the next theorem.

Theorem 20. The necessary and sufficient condition that eqn. (4.6) has a solution is that the ranks of the manifold spanned by the A-vectors, and of that spanned by the A-vectors and vector $\boldsymbol{B}$ together, are equal.

Geometrically speaking, the vector $\boldsymbol{B}$ must lie in the space generated by the vectors $\boldsymbol{A}_i$.

The necessity of this requirement is implied by the equation itself. To demonstrate its sufficiency, let us assume that the condition is fulfilled and that the rank of the column vectors $\boldsymbol{A}_i$ is r. Further, let us also suppose that the equations have been rearranged and the unknowns reordered so that the first r columns and the first r rows of the matrix

$$\mathbf{A} = (\boldsymbol{A}_1\boldsymbol{A}_2 \ldots \boldsymbol{A}_m)$$

are linearly independent (*see* Fig. 3.1, p. 16).

Written out in full, eqn. (4.6) becomes

$$x_1\begin{Bmatrix}a_{1\,1}\\ \cdot\\ \cdot\\ a_{r\,1}\\ a_{r+1\,1}\\ \cdot\\ a_{n\,1}\end{Bmatrix}+\ldots+x_r\begin{Bmatrix}a_{1\,r}\\ \cdot\\ \cdot\\ a_{r\,r}\\ a_{r+1\,r}\\ \cdot\\ a_{n\,r}\end{Bmatrix}+x_{r+1}\begin{Bmatrix}a_{1\,r+1}\\ \cdot\\ \cdot\\ a_{r\,r+1}\\ a_{r+1\,r+1}\\ \cdot\\ a_{n\,r+1}\end{Bmatrix}+\ldots$$

$$+x_m\begin{Bmatrix}a_{1\,m}\\ \cdot\\ \cdot\\ a_{r\,m}\\ a_{r+1\,m}\\ \cdot\\ a_{n\,m}\end{Bmatrix}=\begin{Bmatrix}b_1\\ \cdot\\ \cdot\\ b_r\\ b_{r+1}\\ \cdot\\ b_n\end{Bmatrix}$$

If $n > r$, the last $n - r$ equations are linear combinations of the first r equations; thus any solution of the first r equations will automatically satisfy the remainder. We can therefore ignore the last n–r coordinates in each column vector and write the system of equations as

$$x_1A_1' + \ldots + x_rA_r' + x_{r+1}A_{r+1}' + \ldots + x_mA_m' = B' \qquad (4.7)$$

where the A_i' and B' are r-dimensional vectors.

When $r = m$, this system is identical with the one dealt with in Section 4.1, p. 33, and it therefore has a unique solution which can be determined by Cramer's method.

In the case where $r <$ m, we can assign arbitrary values to the last m–r unknowns, $x_{r+1}, \ldots, x_m$, and remarshal the vector equation as follows—

$$x_1A_1' + \ldots + x_rA_r' = B' - t_1A_{r+1}' \ldots - t_{m-r}A_m'$$

or

$$\mathbf{A}'X' = B' - t_1A_{r+1}' \ldots - t_{m-r}A_m' \qquad . \ (4.8)$$

where $\mathbf{A}'$ is the r-by-r invertible matrix formed from the r independent vectors $A_1', \ldots, A_r'$; X' is an r-dimensional vector with coordinates $x_1, \ldots, x_r$; and $t_1 = x_{r+1}, \ldots, t_{m-r} = x_m$ are m–r arbitrary parameters.

Thus, for any given set of the m–r unknowns, $x_{r+1}, \ldots, x_m$, eqn. (4.8) has the unique solution

$$X' = (\mathbf{A}')^{-1}(B' - t_1A_{r+1}' \ldots - t_{m-r}A_m') \qquad . \ (4.9)$$

To summarize the above results, we can formulate another theorem.

Theorem 21. If B in eqn. (4.6) belongs to the manifold determined by the m vectors A_i, and if the rank of this manifold is r, the equation has a unique solution when $r = m$ and an $(m$–$r)$-fold infinity of solutions (i.e. a solution containing m–r arbitrary parameters) when $r < m$.

If B does not belong to the vector space spanned by the A_i, the equation has no solution.

Exercise 1. Solve the equations

$$\begin{aligned} x + \ y + 3z &= 6\\ 2x + 3y + \ z &= 8\\ 4x + 5y + 7z &= 20\\ x \qquad + 8z &= 10\end{aligned}$$

The *augmented system matrix* $\{A_1A_2A_3B\}$ is

$$\begin{Bmatrix}1 & 1 & 3 & 6\\ 2 & 3 & 1 & 8\\ 4 & 5 & 7 & 20\\ 1 & 0 & 8 & 10\end{Bmatrix}$$

which, by linear operations, reduces to

$$\begin{Bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{Bmatrix}$$

the rank of which is 2 (Section 3.6, p. 22). The first two columns and the first two rows of the system matrix are independent (they are not proportional); thus B belongs to the space spanned by the A_i, and we can ignore the last two equations and substitute a parameter t for the third unknown z. This leads to

$$\begin{Bmatrix}1 & 1\\ 2 & 3\end{Bmatrix}\begin{Bmatrix}x\\ y\end{Bmatrix} = \begin{Bmatrix}6 - 3t\\ 8 - \ t\end{Bmatrix}$$

which when pre-multiplied by

$$\begin{Bmatrix}1 & 1\\ 2 & 3\end{Bmatrix}^{-1} = \begin{Bmatrix}3 & -1\\ -2 & 1\end{Bmatrix}$$

gives us

$$\begin{Bmatrix}x\\ y\end{Bmatrix} = \begin{Bmatrix}3 & -1\\ -2 & 1\end{Bmatrix}\begin{Bmatrix}6 - 3t\\ 8 - \ t\end{Bmatrix} = \begin{Bmatrix}10 - 8t\\ -4 + 5t\end{Bmatrix}$$

The complete solution can now be written

$$\begin{Bmatrix}x\\ y\\ z\end{Bmatrix} = \begin{Bmatrix}10\\ -4\\ 0\end{Bmatrix} + t\,.\begin{Bmatrix}-8\\ 5\\ 1\end{Bmatrix}$$

It comprises one arbitrary parameter (there exists a *single infinity* of solutions to the equations).

Exercise 2. Solve

$$\begin{aligned} 2x \qquad + 2z &= 6 \\ y - 2z &= -3 \\ x - y + 4z &= 7 \\ 5x + 2y + 5z &= 13 \\ -x + 2y + z &= -3 \end{aligned}$$

By linear operations the augmented system matrix reduces to

$$\begin{Bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Bmatrix}$$

the rank of which is 3. The determinant formed by the coefficients of the unknowns in the first three equations is non-zero; hence the $\boldsymbol{A}$-vectors are independent and span a 3-space which contains $\boldsymbol{B}$. We can thus ignore the last two equations, so that the system becomes

$$\begin{Bmatrix} 2 & 0 & 2 \\ 0 & 1 & -2 \\ 1 & -1 & 4 \end{Bmatrix} \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} 6 \\ -3 \\ 7 \end{Bmatrix}$$

$$\begin{Bmatrix} 2 & 0 & 2 \\ 0 & 1 & -2 \\ 1 & -1 & 4 \end{Bmatrix}^{-1} = \tfrac{1}{2} \begin{Bmatrix} 2 & -2 & -2 \\ -2 & 6 & 4 \\ -1 & 2 & 2 \end{Bmatrix}$$

Thus, by pre-multiplication,

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \tfrac{1}{2} \begin{Bmatrix} 2 & -2 & -2 \\ -2 & 6 & 4 \\ -1 & 2 & 2 \end{Bmatrix} \begin{Bmatrix} 6 \\ -3 \\ 7 \end{Bmatrix} = \begin{Bmatrix} 2 \\ -1 \\ 1 \end{Bmatrix}$$

The equations have a unique solution.

Exercise 3. Solve the system of equations

$$\begin{aligned} 2x - 3y + z &= 1 \\ 2x + y \qquad &= 0 \\ 3x + 2y + z &= 1 \\ 6x + 3y + z &= 2 \end{aligned}$$

The augmented system matrix

$$\begin{Bmatrix} 2 & -3 & 1 & 1 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 1 \\ 6 & 3 & 1 & 2 \end{Bmatrix}$$

reduces to

$$\begin{Bmatrix} 0 & 0 & 0 & 1 \\ 0 & 11 & 0 & 0 \\ -9 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{Bmatrix}$$

the rank of which is 4. This immediately shows that $\boldsymbol{B}$ does not lie in the 3-space spanned by the $\boldsymbol{A}_i$.

The equations therefore have no solution (they are said to be contradictory).

Note. The preceding exercises have been presented in such a manner as to enable the reader to visualize geometrically what was being done algebraically. To make the position quite clear, however, we shall give a brief summary of our procedure in purely geometric terms.

$\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3$ and $\boldsymbol{B}$ in Exercise 1 are coplanar, as evidenced by the fact that the rank of the augmented matrix was 2. This permitted us to project the problem onto a plane by ignoring the last two equations (i.e. the components of the vectors along the third and fourth axes). We thus reduced the problem to that of resolving vector $\boldsymbol{B}$ into components along three axes in a plane. This was done by allocating an arbitrary factor t to one (any one) of the $\boldsymbol{A}_i$ (any two of which are independent), and then resolving the combination of this vector and $\boldsymbol{B}$ in terms of the two remaining vectors. Hence, the solution contained one arbitrary parameter.

In Exercise 2 we were required to express a 5-dimensional vector $\boldsymbol{B}$ as a linear combination of three 5-dimensional $\boldsymbol{A}$-vectors. A study of the rank of the system matrix proved that $\boldsymbol{B}$ lay in the 3-dimensional subspace of 5-space spanned by the $\boldsymbol{A}_i$. The problem had a unique solution and could be simplified by shedding two dimensions, thus projecting the vectors into a 3-space.

Exercise 3 posed a problem similar to that treated in Exercise 2, except that in this case $\boldsymbol{B}$ lay outside the 3-dimensional subspace spanned by the $\boldsymbol{A}$-vectors. The equations had, therefore, no solution.

4.3. Nullspace of a Matrix

A special situation arises when $\boldsymbol{B} = \boldsymbol{0}$ and eqn. (4.6), p. 35, becomes

$$x_1\boldsymbol{A}_1 + \ldots + x_m\boldsymbol{A}_m = \boldsymbol{0} \qquad . \quad . \ (4.10)$$

It is obvious that this equation will always have the solution $x_1 = x_2 = \ldots = x_m = 0$.

To obtain the complete solution, we can proceed as described in Section 4.2. When $r = m$, the $\boldsymbol{A}$-vectors are independent and eqn. (4.10) will have only the trivial solution mentioned above. When $r < m$, however, we can assign arbitrary values to the $m - r$ unknowns $x_{r+1}, \ldots, x_m$ and employ eqn. (4.9) by substituting the r-dimensional nullvector $\boldsymbol{0}'$ for $\boldsymbol{B}'$.

This yields the expression

$$\boldsymbol{X}' = -\, t_1(\mathrm{A}')^{-1}\boldsymbol{A}'_{r+1} \ldots - t_{m-r}(\mathrm{A}')^{-1}\boldsymbol{A}'_m \qquad . \ (4.11)$$

from which it is clear that eqn. (4.10) has an $(m - r)$-fold infinity of solutions.

Up to this point we have looked upon the m x_i as coefficients of the m column vectors $\boldsymbol{A}_i$. We shall now interpret these numbers as the coordinates of an m-dimensional vector $\boldsymbol{X}$.

Referring to eqn. (4.11), and bearing in mind that the last $m - r$ coordinates of $\boldsymbol{X}$ are arbitrary parameters, it is readily seen that $\boldsymbol{X}$

is an arbitrary linear combination of $m - r$ vectors of the form

$$R_i = \begin{Bmatrix} -(A')^{-1}A'_{r+i} \\ U_i \end{Bmatrix} \qquad (i = 1, \ldots, m - r) \quad . \ (4.12)$$

where the U_i are the $(m - r)$-dimensional unit vectors.

From the presence of the $m - r$ unit vectors it is obvious that the R_i are linearly independent; this can be seen from the fact that the bottom $(m - r)$-by-$(m - r)$ sub-matrix of the m-by-$(m - r)$ matrix $(R_1R_2 \ldots R_{m-r})$ is a unit matrix the rank of which is $m - r$.

To recapitulate the above results: the solutions of the equation

$$\mathrm{A}X = 0 \qquad . \quad . \quad . \quad . \ (4.13)$$

where A is an n-by-m matrix, X an m-dimensional vector and 0 the n-dimensional nullvector, fill an $(m - r)$-dimensional subspace of m-space, where r is the rank of A. The rank of the nullspace of a square matrix is termed the *nullity* of the matrix.

When $r = m$, eqn. (4.13) has only the trivial solution $X = 0$.

Anticipating the contents of Section 5.2, p. 51, we can say that eqn. (4.13) requires X to be orthogonal to all the row vectors of A.

Exercise 1. Solve the equation

$$\mathrm{A}X = \begin{bmatrix} 2 & 1 & -2 & 7 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ -1 & 0 & 0 & -2 & -1 \\ 2 & 2 & 1 & 5 & 3 \\ 1 & 2 & 1 & 3 & 2 \end{bmatrix} \begin{Bmatrix} x \\ y \\ z \\ u \\ v \end{Bmatrix} = 0$$

By reducing the system matrix we find that $r(\mathrm{A}) = 3$, and by checking the top left-hand 3rd-order subdeterminant of A, it is seen that the two last equations of the set can be ignored. Thus, by treating $u = s$ and $v = t$ as two independent arbitrary parameters we arrive at the equation

$$\begin{bmatrix} 2 & 1 & -2 \\ 0 & 1 & 1 \\ -1 & 0 & 0 \end{bmatrix} \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = s \begin{Bmatrix} -7 \\ 0 \\ 2 \end{Bmatrix} + t \begin{Bmatrix} 0 \\ -1 \\ 1 \end{Bmatrix}$$

which has the solution

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = s \begin{Bmatrix} -2 \\ -1 \\ 1 \end{Bmatrix} + t \begin{Bmatrix} -1 \\ 0 \\ -1 \end{Bmatrix}$$

and the complete solution is

$$\begin{Bmatrix} x \\ y \\ z \\ u \\ v \end{Bmatrix} = s \begin{Bmatrix} -2 \\ -1 \\ 1 \\ 1 \\ 0 \end{Bmatrix} + t \begin{Bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{Bmatrix} = sR + tQ$$

where R and Q span the nullspace of A.

Exercise 2. Discuss the (non-simultaneous) equations

$$2x - y + z + u = 0 \qquad . \quad . \quad . \quad . \ (4.14)$$

and

$$2x - y + z + u = 2 \qquad . \quad . \quad . \quad . \ (4.15)$$

In both equations $m = 4$ and $r = 1$; hence the solutions to eqn. (4.14) will fill a 3-dimensional subspace of 4-space. Following the procedure given in Section 4.2 and in this section, we find that the solution space (the nullspace of the 1-by-4 matrix A) is spanned by the three independent vectors—

$$R_1 = \begin{Bmatrix} \tfrac{1}{2} \\ 1 \\ 0 \\ 0 \end{Bmatrix}, \quad R_2 = \begin{Bmatrix} -\tfrac{1}{2} \\ 0 \\ 1 \\ 0 \end{Bmatrix} \quad \text{and} \quad R_3 = \begin{Bmatrix} -\tfrac{1}{2} \\ 0 \\ 0 \\ 1 \end{Bmatrix}$$

Using eqn. (4.8) we find the solution of eqn. (4.15) to be

$$X = \begin{Bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{Bmatrix} + t_1R_1 + t_2R_2 + t_3R_3 = \begin{Bmatrix} 1 + \tfrac{1}{2}t_1 - \tfrac{1}{2}t_2 - \tfrac{1}{2}t_3 \\ t_1 \\ t_2 \\ t_3 \end{Bmatrix}$$

It should be noted that the complete solution of an equation of the type $\mathrm{A}X = B\ (\neq 0)$ is equal to one solution (any one) of the equation plus the nullspace of A (i.e. the complete solution of the equation $\mathrm{A}X = 0$).

We now define the *complementary space* (nullspace) of a manifold spanned by p n-dimensional vectors, $V_1, \ldots, V_p$, to be the nullspace of the p-by-n matrix

$$\mathrm{A} = \begin{Bmatrix} V_{1t} \\ V_{2t} \\ \cdot \\ \cdot \\ V_{pt} \end{Bmatrix}$$

Together, a manifold and its complementary space fill all n-space.

4.4. Union and Intersection of Vector Spaces

Let two linear vector manifolds in n-space be given as follows—

$$E_p \text{ spanned by } V_1, \ldots, V_p \quad \text{and} \quad E_q \text{ spanned by } W_1, \ldots, W_q$$

The maximum number (s) of linearly independent vectors that can be selected from the two sets is said to span the *union* (join) of the two spaces. The union is the space of lowest possible dimension containing both E_p and E_q.

The nullspace of the union of the nullspaces of E_p and E_q is the *intersection* of E_p and E_q. It is the space of highest possible dimension that can be contained in (is a subspace of) both E_p and E_q.

Taken with a pinch of salt, Fig. 4.1 may prove an aid in visualizing the concepts of union and intersection. The n horizontal divisions

can be taken to represent n independent n-dimensional vectors in terms of which E_p and E_q can be expressed.

If the ranks of E_p and E_q are p and q respectively, the rank of their union is s and that of their intersection t, it is readily seen that

$$p + q = s + t$$

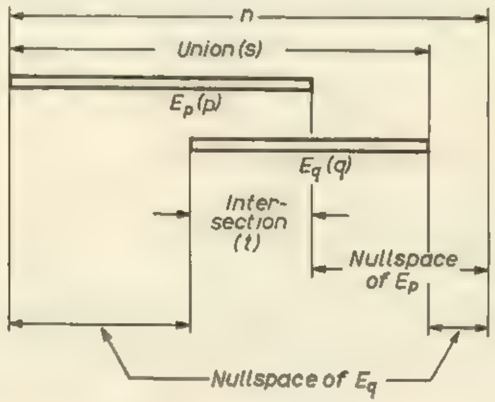


FIG. 4.1. SCHEMATIC REPRESENTATION OF THE CONCEPTS OF UNION AND INTERSECTION

Exercise 1.* Find the union and intersection in an E_5 of an E_3 spanned by

$$V_1 = \begin{Bmatrix} 2 \\ 1 \\ 0 \\ 1 \\ 1 \end{Bmatrix}, \quad V_2 = \begin{Bmatrix} 1 \\ 0 \\ 2 \\ -1 \\ 1 \end{Bmatrix} \quad \text{and} \quad V_3 = \begin{Bmatrix} 2 \\ 1 \\ 1 \\ 0 \\ 0 \end{Bmatrix}$$

and an E_2 spanned by

$$W_1 = \begin{Bmatrix} 3 \\ 2 \\ -1 \\ 2 \\ 0 \end{Bmatrix} \quad \text{and} \quad W_2 = \begin{Bmatrix} 1 \\ 1 \\ 2 \\ 1 \\ 1 \end{Bmatrix}$$

By linear operations the rank of the compound matrix $(V_1 V_2 V_3 W_1 W_2)$ is found to be 4 . V_1, V_2, V_3 and W_2 are shown to be independent, and they can therefore be chosen as a base for the 4-dimensional union of E_3 and E_2.

From the equations

$$\begin{Bmatrix} V_{1t} \\ V_{2t} \\ V_{3t} \end{Bmatrix} X = 0 \quad \text{and} \quad \begin{Bmatrix} W_{1t} \\ W_{2t} \end{Bmatrix} X = 0$$

* In this and in many of the following exercises we shall not describe in detail the steps by which we arrive at a result, but merely outline the method employed or refer to some algebraic technique which has already been demonstrated. Each exercise is thus virtually a problem.

the nullspaces of E_3 and E_2 are found to be spanned by

$$V_4 = \begin{Bmatrix} -1 \\ 1 \\ 1 \\ 1 \\ 0 \end{Bmatrix}, \quad V_5 = \begin{Bmatrix} -3 \\ 5 \\ 1 \\ 0 \\ 1 \end{Bmatrix}$$

and

$$W_3 = \begin{Bmatrix} 5 \\ -7 \\ 1 \\ 0 \\ 0 \end{Bmatrix}, \quad W_4 = \begin{Bmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 0 \end{Bmatrix}, \quad W_5 = \begin{Bmatrix} 2 \\ -3 \\ 0 \\ 0 \\ 1 \end{Bmatrix}$$

respectively.

The rank of matrix $(V_4 V_5 W_3 W_4 W_5)$ is 4, and by computing the corresponding determinant, V_4, V_5, W_4 and W_5 are found to be independent and are selected to span the union of the nullspaces of E_3 and E_2.

To obtain the nullspace of this union (which is the intersection of E_3 and E_2) we could solve the equation

$$\begin{Bmatrix} V_{4t} \\ V_{5t} \\ W_{4t} \\ W_{5t} \end{Bmatrix} X = 0 \quad . \quad . \quad . \quad . \quad . \quad (4.16)$$

but an easier way is to choose the coordinates of X as the cofactors of the elements in the last column of the determinant det $(V_4 V_5 W_4 W_5 0)$. By Theorem 9, p. 20, the vector derived in this manner will satisfy eqn. (4.16). This method yields

$$X = t \begin{Bmatrix} 3 \\ 2 \\ -1 \\ 2 \\ 0 \end{Bmatrix}$$

The intersection of E_3 and E_2 is a straight line.

As a further exercise the reader should sketch a diagram similar to Fig. 4.1.

When working with 5-dimensional vectors, as we did in the previous exercise, we must of necessity lean heavily on our algebra, and can only vaguely visualize the steps we are taking by analogy with the familiar geometry of Euclidean 3-space.

As a contrast to the 5-dimensional problem of Exercise 1, we shall now pose a simple 3-dimensional problem which can readily be pictured. We shall describe our method of solution in words and give the calculated results, but the actual calculations are left to the reader.

Exercise 2. The problem is to find the union and intersection of E_2 and E_2' spanned by

$$V_1 = \begin{Bmatrix} 2 \\ 1 \\ 1 \end{Bmatrix}, \quad V_2 = \begin{Bmatrix} 0 \\ 2 \\ 1 \end{Bmatrix}; \quad \text{and} \quad W_1 = \begin{Bmatrix} 1 \\ 0 \\ 1 \end{Bmatrix}, \quad W_2 = \begin{Bmatrix} 0 \\ 3 \\ 4 \end{Bmatrix}$$

respectively.

By checking the appropriate determinants we can see that no three of the vectors V_1, V_2, W_1 and W_2 are coplanar; their union is thus the whole of 3-space. V_1 and V_2 are independent and contain a plane P; their complement is a line L, perpendicular to P, and determined by the vector

$$V_3 = \begin{Bmatrix} -1 \\ -2 \\ 4 \end{Bmatrix}$$

Likewise, W_1 and W_2 span a plane P' and their complement is a line L', perpendicular to P', and lying along the vector

$$W_3 = \begin{Bmatrix} 3 \\ 4 \\ -3 \end{Bmatrix}$$

The union of the nullspaces of E_2 and E_2' is a plane through L and L'. The complement of this space is a line L'', orthogonal to both L and L', and determined by

$$Q = \begin{Bmatrix} -10 \\ 9 \\ 2 \end{Bmatrix}$$

L'' is at right angles to L and therefore lies in P; by a similar argument it must also lie in P'. Hence L'' is the line along which P and P' intersect.

4.5. Non-Centred Vector Spaces

The vector spaces introduced and defined in Section 2.5, p. 11, are said to be centred because they contain the origin $\boldsymbol{0}$.

A vector space defined by the equation

$$X = A_0 + t_1A_1 + \dots t_mA_m \qquad (4.17)$$

where the A_i $(i = 0, 1, \dots, m)$ are n-dimensional vectors, and $t_1, \dots, t_m$ independent arbitrary parameters, is called *non-centred* when it does not include $\boldsymbol{0}$.

Obviously, the sufficient condition that the space defined by eqn. (4.17) shall not include the origin is that the set of vectors $A_0, A_1, \dots, A_m$ are linearly independent. Hence $n > m$.

Suppose we have two non-centred spaces, E_p and E_q, in n-space and wish to determine their intersection.

Let E_p be defined by

$$X = P_0 + t_1P_1 + \dots + t_pP_p \qquad (4.18)$$

and E_q by

$$X = Q_0 + s_1Q_1 + \dots + s_qQ_q \qquad (4.19)$$

where each set of base vectors is independent and the t's and s's are independent arbitrary parameters.

A point of intersection of E_p and E_q will lie simultaneously in both spaces, i.e.

$$P_0 + t_1P_1 + \dots + t_pP_p = Q_0 + s_1Q_1 + \dots + s_qQ_q$$

$$\text{or} \quad t_1P_1 + \dots + t_pP_p - s_1Q_1 - \dots - s_qQ_q = Q_0 - P_0 \qquad (4.20)$$

This equation can be discussed and solved by the method described in Section 4.2, p. 35. Generally the solutions will fill a space whose dimension will not exceed either p or q (whichever is the smaller) and never be less than $p + q - n$.

When $p + q < n$, E_p may miss each other altogether, i.e. have no points in common. Compare in this connexion two E_1's in 3-space, at least one of which is non-centred (two straight lines which do not both pass through the origin).

Exercise 1. Prove that an E_m and an E_{n-m} in n-space generally intersect in an E_0 (a point).

If the two spaces have no directions in common the set of vectors consisting of their $m + (n - m) = n$ base vectors will be linearly independent and eqn. (4.20) will yield a unique solution which will determine the point of intersection.

Exercise 2. Discuss the intersection in 4-space of the non-centred space

$$X = P_0 + t_1P_1 + t_2P_2 = \begin{Bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{Bmatrix} + t_1 \begin{Bmatrix} 0 \\ -1 \\ 2 \\ 2 \end{Bmatrix} + t_2 \begin{Bmatrix} 2 \\ 0 \\ -1 \\ 1 \end{Bmatrix}$$

and the centred space

$$X = s_1Q_1 = s_1 \begin{Bmatrix} 2 \\ -1 \\ 1 \\ 3 \end{Bmatrix}$$

By means of linear column operations we find $r(Q_1P_0P_1P_2) = 3$, and $r(Q_1P_1P_2) = 2$. Thus the two spaces have no point in common.

4.6. Right and Left Inverses of Singular Matrices

In Section 3.9, p. 28, it was shown that an n-by-n matrix $\mathbf{A}$ of rank n has a unique inverse $\mathbf{A}^{-1}$ such that

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

We shall consider here in greater detail why a singular square matrix cannot have an inverse.

Let $\mathbf{A}$ be an n-by-n matrix of rank $r < n$ and let us seek a postfactor $\mathbf{X}$ such that

$$\mathbf{AX} = \mathbf{I} \qquad (4.21)$$

It follows from the shape of the matrices **A** and **I** that **X**, if it exists, must also have n rows and n columns.

By considering **X** and **I** as consisting of column vectors, we have

$$\mathbf{X} = (X_1 \ldots X_i \ldots X_n)$$

and

$$\mathbf{I} = (U_1 \ldots U_i \ldots U_n)$$

where U_i is an n-dimensional unit vector with its unit element in the ith place.

Eqn. (4.21) can now be broken down into n sets of n linear equations with n unknowns—

$$\mathbf{A}X_i = U_i \quad . \quad . \quad . \quad . \quad (4.22)$$

Since $r(\mathbf{A}) = r\ (< n)$, the column vectors of **A** will span an r-dimensional subspace of n-space. Thus, for at least $n - r$ values of i, the vectors U_i will lie outside this subspace and the corresponding eqns. (4.22) will have no solution. Hence, eqn. (4.21) has no solution.

An exactly analogous train of reasoning employing row vectors can be pursued to prove the non-existence of a left inverse of **A**.

Another type of singular matrix is a rectangular matrix with more rows than columns or more columns than rows.

Let **A** be an m-by-n matrix with maximum possible rank $r(\mathbf{A}) = m$ or n (whichever is the smaller).

If we wish to define a right inverse of **A** according to eqn. (4.21), **X** will have to be an n-by-m matrix, which in turn makes **I** the m-by-m unit matrix. As before, we calculate **X** column by column by means of eqn. (4.22) which when stated in greater detail reads—

$$x_{1i}A_1 + x_{2i}A_2 + \ldots + x_{ni}A_n = U_i \quad . \quad . \quad (4.23)$$

where the A's and U_i are m-dimensional vectors.

When $n < m$, at least $m - n$ of the unit vectors are not contained in the n-dimensional subspace of m-space spanned by the A-vectors, and therefore at least $m - n$ of the m eqns. (4.23) are unsolvable. Thus for $n < m$ no inverse exists.

In cases where $n > m$, the A-vectors span the whole m-space ($r(\mathbf{A}) = m$) and each of the m eqns. (4.23) has an $(n - m)$-fold infinity of solutions. Thus an $m(n - m)$-fold infinity of inverses exists.

Exercise. Calculate the right inverse of A, where

$$\mathbf{A} = \begin{Bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{Bmatrix}$$

Let

$$\mathbf{X} = \begin{Bmatrix} x_1 & x_2 \\ y_1 & y_2 \\ z_1 & z_2 \end{Bmatrix}$$

To determine column 1 of **X** we must solve equation

$$\begin{Bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{Bmatrix} \begin{Bmatrix} x_1 \\ y_1 \\ z_1 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}$$

which is similar to the equation discussed in Exercise 1, Section 4.2.

The solution, which we leave it to the reader to work out in detail, is

$$X_1 = \begin{Bmatrix} 1 + s \\ -1 - 2s \\ s \end{Bmatrix}$$

Similarly, the equation for the second column of **X** has the solution

$$X_2 = \begin{Bmatrix} t \\ 1 - 2t \\ t \end{Bmatrix}$$

The right inverse of A is thus

$$\mathbf{X} = \begin{Bmatrix} 1 + s & t \\ -1 - 2s & 1 - 2t \\ s & t \end{Bmatrix}$$

4.7. The Factorized Inverse

In this section we shall discuss a method of inverting large matrices which is fundamentally important in the field of diakoptics.

Let the following matrix equation be given—

$$\mathbf{A}X = F \quad . \quad . \quad . \quad . \quad (4.24)$$

where **A** is a large n-by-n matrix of rank n, and suppose that it is possible to decompose **A** as follows—

$$\mathbf{A} = \mathbf{B} + \mathbf{KPK}_t \quad . \quad . \quad . \quad (4.25)$$

where **B** is non-singular and easily invertible, **P** is a p-by-p non-singular matrix, and **K** is an n-by-p matrix of rank $p\ (< n)$.

Eqn. (4.24) now becomes

$$\mathbf{B}X + \mathbf{KPK}_tX = F \quad . \quad . \quad . \quad (4.26)$$

or (since $\mathbf{B}^{-1}$ exists)

$$X = \mathbf{B}^{-1}F - \mathbf{B}^{-1}\mathbf{KPK}_tX \quad . \quad . \quad (4.27)$$

Pre-multiplying this equation by $\mathbf{K}_t$,

$$\mathbf{K}_tX = \mathbf{K}_t\mathbf{B}^{-1}F - \mathbf{K}_t\mathbf{B}^{-1}\mathbf{KPK}_tX \quad . \quad . \quad (4.28)$$

Rearranging this equation and solving for $\mathbf{K}_tX$,

$$\mathbf{K}_tX = (\mathbf{I} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{KP})^{-1}\mathbf{K}_t\mathbf{B}^{-1}F \quad . \quad . \quad (4.29)$$

Substituting (4.29) in (4.27), we find that

$$\begin{aligned} X &= [\mathbf{I} - \mathbf{B}^{-1}\mathbf{KP}(\mathbf{I} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{KP})^{-1}\mathbf{K}_t]\mathbf{B}^{-1}F \\ &= [\mathbf{I} - \mathbf{B}^{-1}\mathbf{K}(\mathbf{P}^{-1} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{K})^{-1}\mathbf{K}_t]\mathbf{B}^{-1}F \quad . \quad (4.30) \end{aligned}$$

By comparison with (4.26) it is clear that

$$\mathbf{A}^{-1} = [\mathbf{I} - \mathbf{B}^{-1}\mathbf{K}(\mathbf{P}^{-1} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{K})^{-1}\mathbf{K}_t]\mathbf{B}^{-1} \quad . \quad . \quad (4.31)$$

Apart from the addition and multiplication of matrices, the inversion of **A** has been reduced to the inversion of two other matrices: **B**, which by hypothesis is easily inverted, and **P** which is appreciably smaller than **A** ($p < n$).

Before the validity of eqn. (4.31) can be accepted, however, we must show that the inversion of $\mathbf{I} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{KP}$ (eqn. (4.30)) is permissible, i.e. that $r(\mathbf{I} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{KP}) = p$. Post-multiplication by $\mathbf{K}_t$ will not alter the rank of the matrix because $r(\mathbf{K}) = p$. Hence

$$\begin{aligned} \mathbf{K}_t + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{KPK}_t &= \mathbf{K}_t(\mathbf{I} + \mathbf{B}^{-1}\mathbf{KPK}_t) \\ &= \mathbf{K}_t(\mathbf{I} + \mathbf{B}^{-1}(\mathbf{A} - \mathbf{B})) \\ &= \mathbf{K}_t\mathbf{B}^{-1}\mathbf{A} \, . \quad . \quad . \quad . \quad (4.32) \end{aligned}$$

the rank of which is p, since $r(\mathbf{B}^{-1}\mathbf{A}) = n$ (*see* the Exercise on p. 27).

Exercise. Prove by direct multiplication that eqn. (4.31) is correct. Post-multiplication of **A** by this equation gives

$$\begin{aligned} \mathbf{AA}^{-1} &= (\mathbf{B} + \mathbf{KPK}_t)[\mathbf{I} - \mathbf{B}^{-1}\mathbf{K}(\mathbf{P}^{-1} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{K})^{-1}\mathbf{K}_t]\mathbf{B}^{-1} \\ &= [\mathbf{B} - \mathbf{K}(\mathbf{P}^{-1} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{K})^{-1}\mathbf{K}_t + \mathbf{KPK}_t - \mathbf{KPK}_t\mathbf{B}^{-1}\mathbf{K}(\mathbf{P}^{-1} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{K})^{-1}\mathbf{K}_t]\mathbf{B}^{-1} \\ &= [\mathbf{B} + \mathbf{KPK}_t - \mathbf{KP}(\mathbf{P}^{-1} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{K})(\mathbf{P}^{-1} + \mathbf{K}_t\mathbf{B}^{-1}\mathbf{K})^{-1}\mathbf{K}_t]\mathbf{B}^{-1} \\ &= (\mathbf{B} + \mathbf{KPK}_t - \mathbf{KPK}_t)\mathbf{B}^{-1} = \mathbf{I} \quad \text{(Q.E.D.)} \end{aligned}$$

As a further exercise the reader should check that the product $\mathbf{A}^{-1}\mathbf{A}$ is also equal to the unit matrix **I**.

PROBLEMS

4.1. Discuss the simultaneous linear equations

$$\begin{aligned} 2x + y - z + u &= 6 \\ x + y + z + 4u &= 6 \end{aligned}$$

4.2. Determine a such that the simultaneous linear equations

$$\begin{aligned} x + 2y - z &= a \\ -x + y + z &= -2 \\ 2x + 2z &= 6 \\ x - y + z &= 4 \\ -x + y &= -3 \end{aligned}$$

have a unique solution.

4.3. Discuss the simultaneous linear equations

$$\begin{aligned} x - y &= 5 \\ 2x + y &= 4 \\ -x + 2y &= -7 \\ x - 3y &= 9 \\ 2x + 2y &= 2 \end{aligned}$$

4.4. Determine the constants a and b such that the simultaneous linear equations

$$\begin{aligned} x + y + z &= -1 \\ -x + 2z &= a \\ x - 2y - z &= -3 \\ 2x + 3y + z &= b \\ -x + 2y - z &= 7 \end{aligned}$$

have a solution.

4.5. Determine the nullspaces of the matrices

$$\mathbf{A} = \begin{Bmatrix} 1 & 1 & 2 \\ 0 & 2 & -1 \\ 1 & -1 & 1 \end{Bmatrix}, \quad \mathbf{B} = \begin{Bmatrix} -1 & 2 & 3 \\ 2 & 1 & 4 \\ 0 & 1 & 2 \end{Bmatrix}, \quad \mathbf{C} = \begin{Bmatrix} 1 & 2 & -3 \\ 2 & 4 & -6 \\ -1 & -2 & 3 \end{Bmatrix}$$

4.6. Solve the equation $\mathbf{A}X = \mathbf{0}$, where

$$\mathbf{A} = \begin{Bmatrix} 0 & 1 & 1 & 0 & 1 \\ 2 & -1 & 2 & 2 & -1 \\ 1 & 2 & 3 & 1 & 1 \\ 2 & 0 & 3 & 2 & 0 \\ 0 & 5 & 4 & 0 & 3 \end{Bmatrix}$$

4.7. Check that $(X_0)_t = (1 \;\; 0 \;\; -1 \;\; 2 \;\; 1)$ is a solution of the equation $\mathbf{A}X = B$, where $B_t = (0 \;\; 3 \;\; 1 \;\; 3 \;\; -1)$ and A is the 5-by-5 matrix given in Problem 4.6. Give the complete solution.

4.8. Determine the intersection of the two manifolds

$$X_1 = A_0 + t_1A_1 + t_2A_2$$

and

$$X_2 = s_1B_1$$

where $A_0 = \begin{Bmatrix} 2 \\ -5 \\ 3 \end{Bmatrix}, \quad A_1 = \begin{Bmatrix} 1 \\ -2 \\ 1 \end{Bmatrix}, \quad A_2 = \begin{Bmatrix} 2 \\ 1 \\ 1 \end{Bmatrix}, \quad B_1 = \begin{Bmatrix} 0 \\ 1 \\ -1 \end{Bmatrix}$

4.9. Show that, if an n-by-n matrix has the rank r and the nullity s, then $r + s = n$.

4.10. Prove that a linear manifold (vector space) and its nullspace (complementary space) fill all n-space.

4.11. Apply the method employed in the discussion of the non-existence of the inverse of a singular matrix given in Section 4.6, p. 45, to the matrix

$$\begin{Bmatrix} 1 & 2 & 0 \\ -1 & 1 & -3 \\ 2 & 1 & 3 \end{Bmatrix}$$

4.12. Calculate the left inverse of

$$A = \begin{Bmatrix} 2 & 1 \\ 1 & 1 \\ -1 & 0 \\ 1 & -1 \end{Bmatrix}$$

4.13. Let a non-centred 2-dimensional subspace E_2 of 5-space be determined (as in Section 4.5, p. 44) by the vectors

$$A_0 = \begin{Bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 1 \end{Bmatrix}, \quad A_1 = \begin{Bmatrix} 1 \\ 2 \\ -1 \\ 1 \\ 0 \end{Bmatrix} \quad \text{and} \quad A_2 = \begin{Bmatrix} -1 \\ 0 \\ 1 \\ -1 \\ 2 \end{Bmatrix}$$

Determine a 2-dimensional space through the origin which does not intersect E_2.

CHAPTER 5

Orthonormal Matrices—Compound Matrices

5.1. Outer Multiplication of Vectors

FROM the manner in which matrix multiplication was defined it follows that two like (equi-dimensional) vectors can be compounded in two different ways.

Outer (or dyadic) multiplication of two vectors A and B is written

$$AB_t = \begin{Bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_n \end{Bmatrix} \{b_1\, b_2 \, . \, . \, . \, b_n\} = \begin{Bmatrix} a_1b_1 & . \, . \, . & a_1b_n \\ a_2b_1 & . \, . \, . & a_2b_n \\ \cdot & & \cdot \, \cdot \, \cdot \, \cdot \\ \cdot & & \cdot \, \cdot \, \cdot \, \cdot \\ a_nb_1 & . \, . \, . & a_nb_n \end{Bmatrix}$$

The result is a square matrix the rank of which is only 1, however, since all the columns (and all the rows) are proportional.

5.2. Scalar Multiplication of Vectors

When two like vectors are multiplied as follows—

$$A_tB = a_1b_1 + a_2b_2 + \, . \, . \, . + a_nb_n$$

the result is a 1-by-1 matrix or scalar. The product is called the *inner* or scalar product of A and B. When $A_tB = 0$, A and B are said to be orthogonal.

If the scalar product of a vector by itself is unity, we say that the vector is *normalized* (the vector is of unit length)—

$$A_tA = a_1^2 + a_2^2 + \, . \, . \, . + a_n^2 = 1$$

Any non-zero vector can be normalized by multiplication by $1/\sqrt{(A_tA)}$.

Exercise 1. In plane Cartesian coordinates all unit vectors can be written in terms of a parameter v as

$$A = \begin{Bmatrix} \cos v \\ \sin v \end{Bmatrix}$$

where $0 \leqslant v \leqslant 2\pi$. For all values of v the vector

$$B = \begin{Bmatrix} -\sin v \\ \cos v \end{Bmatrix}$$

will also be normalized and orthogonal to A.

If any two of a set of vectors are orthogonal, the vectors are said said to be *mutually orthogonal.*

Theorem 22. A set of mutually orthogonal vectors are linearly independent.

To prove this theorem let us assume that the set of vectors $A_1, A_2, \ldots, A_q$ are normalized and mutually orthogonal. From the equation

$$k_1A_1 + k_2A_2 + \ldots + k_qA_q = 0$$

we derive (by pre-multiplication by A_{1t})

$$k_1A_{1t}A_1 + k_2A_{1t}A_2 + \ldots + k_qA_{1t}A_q$$
$$= k_1 . 1 + k_2 . 0 + \ldots + k_q . 0 = 0$$

Hence $k_1 = 0$, and similarly $k_2 = k_3 = \ldots = k_q = 0$, which proves that the vectors are independent.

From the above theorem it is clear that a set of mutually orthogonal n-dimensional vectors cannot number more than n. The unit vectors $U_1, U_2, \ldots, U_n$ are an example of such a set.

Theorem 23. A set of q n-dimensional mutually orthogonal vectors ($q < n$) can always be augmented by $n - q$ vectors to form an orthogonal set of n vectors.

Let $A_1, \ldots, A_q$ be q orthogonal vectors. An additional vector A_{q+1} must satisfy the q equations $A_{1t}A_{q+1} = 0, \ldots, A_{qt}A_{q+1} = 0$ with n unknowns (the coordinates of A_{q+1}). These equations can be written

$$\begin{Bmatrix} A_{1t} \\ A_{2t} \\ \cdot \\ \cdot \\ A_{qt} \end{Bmatrix} A_{q+1} = \mathrm{A}A_{q+1} = 0 \quad . \quad . \quad . \ (5.1)$$

The q A-vectors are mutually orthogonal; thus $r(\mathrm{A}) = q$, and eqn. (5.1) has an $(n - q)$-fold infinity of solutions from which A_{q+1} can be selected. By induction the theorem is seen to hold in all cases where $q < n$.

As a corollary to Theorem 23 the following statement is readily seen to be true: In a p-dimensional vector space ($p > 1$) it is always possible to select a set of p mutually orthogonal vectors.

The first vector, A_1, of the set can be chosen arbitrarily. The second vector, A_2, must lie in the complementary space of A_1 (a $(p - 1)$-space). To determine a third vector we can choose any vector in the $(p - 2)$-dimensional complement of the manifold spanned by A_1 and A_2. This process can be continued until p vectors have been found. When determining the last vector of the set A_p, we have no longer any choice of direction since A_p must lie on the unique line forming the nullspace of the set $A_1, \ldots, A_{p-1}$.

Exercise 2. Calculate a set of orthogonal 3-dimensional vectors.

We have complete freedom in selecting the first vector of the set, so let us take

$$A_1 = \begin{Bmatrix} 2 \\ 1 \\ 2 \end{Bmatrix}$$

The coordinates x, y, z of the second vector must satisfy the equation

$$2x + y + 2z = 0$$

which is the equation for a plane through the origin and normal to A_1. Suppose we take $x = 0, y = -2$ and $z = 1$. Then

$$A_2 = \begin{Bmatrix} 0 \\ -2 \\ 1 \end{Bmatrix}$$

A_1 and A_2 span a plane and A_3 must therefore lie along the line through the origin and perpendicular to the plane; the only choice left to us is that of the magnitude and orientation of A_3. By means of the method used in Exercise 1, Section 4.4, to calculate the intersection of an E_3 and an E_2 (which in the 3-dimensional case becomes identical with the formula used to compute the cross-product of two vectors), we find that

$$A_3 = \begin{Bmatrix} 5 \\ -2 \\ -4 \end{Bmatrix}$$

5.3. Orthonormal Matrices

A square matrix in which the column vectors are normalized and mutually orthogonal is termed an *orthonormal matrix.*

Let $A_1, \ldots, A_n$ be the column vectors of A. The condition for orthonormality is

$$A_{rt}A_s = a_{1r}a_{1s} + \ldots + a_{nr}a_{ns} = \begin{cases} 1 & \text{when} \quad r = s \\ 0 & \text{when} \quad r \neq s \end{cases}$$

These n^2 formulae can be written concisely as

$$\mathrm{A}_t\mathrm{A} = \mathrm{I} \quad . \quad . \quad . \quad . \ (5.2)$$

Thus the transpose of an orthonormal matrix is identical with its inverse. Since det (A_t) = det (A), it follows the det (A) = +1 or −1.

Also, because a square matrix and its inverse commute, $\mathbf{AA}_t = \mathbf{I}$, so that the row vectors of **A** are also normalized and mutually orthogonal.

A few important properties of orthonormal matrices are readily verified—

(i) If **A** and **B** are orthonormal, their product **AB** is also orthonormal.

Proof. $(\mathbf{AB})_t(\mathbf{AB}) = \mathbf{B}_t\mathbf{A}_t\mathbf{AB} = \mathbf{B}_t\mathbf{IB} = \mathbf{B}_t\mathbf{B} = \mathbf{I}$ (Q.E.D.).

(ii) Multiplication of orthonormal matrices is associative. This follows immediately from the general associativity of matrix multiplication.

(iii) The unit matrix is orthonormal.

Proof. $\mathbf{I}_t\mathbf{I} = \mathbf{II} = \mathbf{I}$.

(iv) The inverse of an orthonormal matrix is itself orthonormal.

Proof. $(\mathbf{A}^{-1})_t\mathbf{A}^{-1} = (\mathbf{A}_t)^{-1}\mathbf{A}^{-1} = (\mathbf{AA}_t)^{-1} = \mathbf{I}^{-1} = \mathbf{I}$.

Our reason for including the apparently unnecessary property (ii) will become clear when we define a mathematical group (Section 6.5).

Exercise 1. In Exercise 2, Section 5.2, we selected three orthogonal vectors in 3-space. The compound matrix $A = (A_1A_2A_3)$ formed from these vectors is orthogonal but not orthonormal. The reader should check the products $\mathbf{AA}_t$ and $\mathbf{A}_t\mathbf{A}$ (the second is a diagonal matrix).

A can be converted into an orthonormal matrix by normalizing its column vectors. This results in

$$\mathbf{A}' = \begin{Bmatrix} 2/3 & 0 & 5/(3\sqrt{5}) \\ 1/3 & -2/\sqrt{5} & -2/(3\sqrt{5}) \\ 2/3 & 1/\sqrt{5} & -4/(3\sqrt{5}) \end{Bmatrix}$$

Note that the row vectors of **A**′ are unit vectors and mutually orthogonal (which the row vectors of **A** were not).

Exercise 2. In 2-space all orthonormal matrices are given parametrically by

$$\mathbf{A} = \begin{Bmatrix} \cos v & \mp \sin v \\ \sin v & \pm \cos v \end{Bmatrix}$$

5.4. Partitioned Matrices

On a number of occasions we have combined a set of vectors to form a matrix. Conversely, it is feasible to partition a matrix into submatrices. However, when partitioning the factors of a matrix product, care must be taken to ensure that the partial products that result from the process are defined.

A little experimentation will show that the rows of the pre-factor and the columns of the post-factor can be partitioned *ad lib.*, whereas the columns of the pre-factor and the rows of the post-factor must be subdivided according to the same pattern.

This principle is best illustrated by means of a sketch (Fig. 5.1).

The two matrices to be multiplied are indicated by rectangles. If their product is defined, the number of columns in the pre-factor is equal to the number of rows in the post-factor. This is the only condition governing the "shape" of the factors.

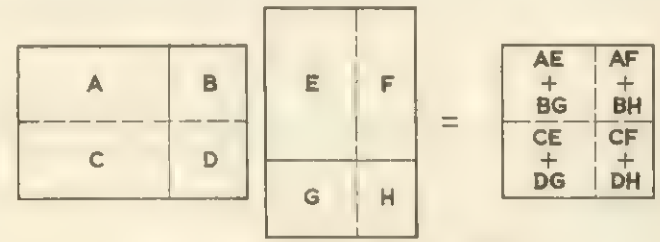


FIG. 5.1. PARTITIONING A MATRIX PRODUCT

We can now subdivide the matrices vertically and horizontally, as shown by the full and broken lines, and multiply them together as though the blocks into which they have been partitioned were scalar elements instead of submatrices. The position and number of the broken lines are not limited in any way, but the position of the full lines is subject to the condition that the partial products **AE**, **BG**, **CF**, **DH**, etc., shall be defined.

It is easily verified by inspection that the matrix arrived at by partitioned multiplication is identical with the one obtained in the ordinary way.

Exercise. The inverse of a partitioned matrix of the type

$$\mathbf{A} = \begin{Bmatrix} \mathbf{K} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{L} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{M} \end{Bmatrix}$$

where **K**, **L** and **M** are square but not necessarily like matrices, is

$$\mathbf{A}^{-1} = \begin{Bmatrix} \mathbf{K}^{-1} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{L}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{M}^{-1} \end{Bmatrix}$$

Compare this result with that of Exercise 1, Section 3.9, and note also that det (**A**) = det (**K**) det (**L**) det (**M**). Thus $\mathbf{A}^{-1}$ is defined only if all three determinants of the square submatrices are non-zero.

5.5. Partitioned Inverses

One aspect of the method of subdivision, the *partitioned inverse*, is important enough to warrant a separate section.

Suppose we have two non-singular matrices $\mathbf{Z}$ and $\mathbf{Y}$ that are each other's inverses; i.e.

$$\mathbf{ZY} = \mathbf{I} = \mathbf{YZ} \quad . \quad . \quad . \quad . \quad (5.3)$$

Furthermore, let $\mathbf{Z}$ be given and $\mathbf{Y}$ unknown. By means of partitioning it is now possible to find $\mathbf{Y}$ without having to determine $\mathbf{Z}^{-1}$ explicitly. This is done by partitioning $\mathbf{Z}$ and $\mathbf{Y}$ as follows—

$$\begin{Bmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 \\ \mathbf{Z}_3 & \mathbf{Z}_4 \end{Bmatrix} \begin{Bmatrix} \mathbf{Y}_1 & \mathbf{Y}_2 \\ \mathbf{Y}_3 & \mathbf{Y}_4 \end{Bmatrix} = \begin{Bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{Bmatrix} \quad . \quad . \quad (5.4)$$

where $\mathbf{Z}_1$, $\mathbf{Z}_4$, $\mathbf{Y}_1$ and $\mathbf{Y}_4$ are square and the two unit matrices in the right-hand member of the equation are also square but not necessarily like.

Multiplying the two matrix rows of $\mathbf{Z}$ by the first matrix column of $\mathbf{Y}$ yields the equations

$$\mathbf{Z}_1\mathbf{Y}_1 + \mathbf{Z}_2\mathbf{Y}_3 = \mathbf{I} \quad . \quad . \quad . \quad (5.5)$$

and

$$\mathbf{Z}_3\mathbf{Y}_1 + \mathbf{Z}_4\mathbf{Y}_3 = \mathbf{0} \quad . \quad . \quad . \quad (5.6)$$

Solving this equation for $\mathbf{Y}_3$ gives

$$\mathbf{Y}_3 = -\mathbf{Z}_4^{-1}\mathbf{Z}_3\mathbf{Y}_1 \quad . \quad . \quad . \quad (5.7)$$

and, substituting this in eqn. (5.5), and solving for $\mathbf{Y}_1$, we find that

$$\mathbf{Y}_1 = (\mathbf{Z}_1 - \mathbf{Z}_2\mathbf{Z}_4^{-1}\mathbf{Z}_3)^{-1} \quad . \quad . \quad . \quad (5.8)$$

and, from eqn. (5.7),

$$\mathbf{Y}_3 = -\mathbf{Z}_4^{-1}\mathbf{Z}_3(\mathbf{Z}_1 - \mathbf{Z}_2\mathbf{Z}_4^{-1}\mathbf{Z}_3)^{-1} \quad . \quad . \quad (5.9)$$

In order to find $\mathbf{Y}_2$ and $\mathbf{Y}_4$ in terms of the matrix elements of $\mathbf{Z}$, we could multiply the matrix rows of $\mathbf{Z}$ by the second matrix column of $\mathbf{Y}$. Such an approach leads to formulae which can be derived immediately from eqns. (5.8) and (5.9) by interchanging the indices 1 and 4, and 2 and 3.

We shall, however, follow another path by employing the fact that $\mathbf{Z}$ and $\mathbf{Y}$ commute. From the identity $\mathbf{YZ} = \mathbf{I}$ we get

$$\mathbf{Y}_1\mathbf{Z}_2 + \mathbf{Y}_2\mathbf{Z}_4 = 0 \quad . \quad . \quad . \quad (5.10)$$

which yields

$$\mathbf{Y}_2 = -\mathbf{Y}_1\mathbf{Z}_2\mathbf{Z}_4^{-1} = -(\mathbf{Z}_1 - \mathbf{Z}_2\mathbf{Z}_4^{-1}\mathbf{Z}_3)^{-1}\mathbf{Z}_2\mathbf{Z}_4^{-1} \quad . \quad (5.11)$$

and

$$\mathbf{Y}_3\mathbf{Z}_2 + \mathbf{Y}_4\mathbf{Z}_4 = \mathbf{I} \quad . \quad . \quad . \quad (5.12)$$

which finally gives us

$$\mathbf{Y}_4 = (\mathbf{I} - \mathbf{Y}_3\mathbf{Z}_2)\mathbf{Z}_4^{-1} = (\mathbf{I} + \mathbf{Z}_4^{-1}\mathbf{Z}_3(\mathbf{Z}_1 - \mathbf{Z}_2\mathbf{Z}_4^{-1}\mathbf{Z}_3)^{-1}\mathbf{Z}_2)\mathbf{Z}_4^{-1} \quad . \quad . \quad . \quad (5.13)$$

Exercise. Invert the matrix

$$\mathbf{Z} = \begin{Bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 0 & 3 \\ 1 & 2 & -1 & 1 \\ 0 & 1 & 0 & 2 \end{Bmatrix}$$

by partitioning.

We shall subdivide the matrix into four 2-by-2 submatrices. The first matrix to be computed is

$$\mathbf{Z}_4^{-1} = \begin{Bmatrix} -1 & 1 \\ 0 & 2 \end{Bmatrix}^{-1} = \tfrac{1}{2}\begin{Bmatrix} -2 & 1 \\ 0 & 1 \end{Bmatrix}$$

Next we compute

$$\mathbf{Y}_1 = (\mathbf{Z}_1 - \mathbf{Z}_2\mathbf{Z}_4^{-1}\mathbf{Z}_3)^{-1} = \begin{Bmatrix} 1 & 3 \\ 0 & -2 \end{Bmatrix}$$

$$\mathbf{Y}_2 = -\mathbf{Y}_1\mathbf{Z}_2\mathbf{Z}_4^{-1} = \begin{Bmatrix} 0 & -5 \\ 0 & 3 \end{Bmatrix}$$

$$\mathbf{Y}_3 = -\mathbf{Z}_4^{-1}\mathbf{Z}_3\mathbf{Y}_1 = \begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}$$

$$\mathbf{Y}_4 = (\mathbf{I} - \mathbf{Y}_3\mathbf{Z}_2)\mathbf{Z}_4^{-1} = \begin{Bmatrix} -1 & 0 \\ 0 & -1 \end{Bmatrix}$$

The actual calculations, which are quite straightforward, are left to the reader.

Finally,

$$\mathbf{Y} = \begin{Bmatrix} 1 & 3 & 0 & -5 \\ 0 & -2 & 0 & 3 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{Bmatrix}$$

A check will show that the product of $\mathbf{Z}$ and $\mathbf{Y}$ is equal to the unit matrix.

PROBLEMS

5.1. Add a third vector A_3 to the set of vectors A_1 and A_2 with coordinates (2, 1, 2) and (1, 0, −1), respectively, so as to form an orthogonal system. Construct an orthonormal matrix from the three vectors.

5.2. Study the properties of the square matrix $\mathbf{U} = (U_nU_{n-1} \ldots U_2U_1)$, where the U_i $(i = 1, 2, \ldots, n)$ are the n n-dimensional unit vectors. Is $\mathbf{U}$ orthonormal?

5.3. Investigate the properties of an n-by-n matrix, $\mathbf{P}$, formed by permuting the rows of the n-by-n unit matrix $\mathbf{I}$. Show that $\mathbf{P}$ is orthonormal and that pre-multiplication of a square matrix, $\mathbf{A}$, by $\mathbf{P}$ permutes the rows of $\mathbf{A}$ in the same manner in which the rows of $\mathbf{I}$ were permuted to form $\mathbf{P}$.

5.4. Study the properties of $\mathbf{P}$ as a post-factor.

5.5. Demonstrate that the matrices

$$\mathbf{A} = \begin{Bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{Bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{Bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{Bmatrix}$$

are orthonormal and show that $\mathbf{AB} \neq \mathbf{BA}$.

5.6. Show that $\mathbf{A}^4 = \mathbf{B}^4 = (\mathbf{AB})^3 = (\mathbf{BA})^3 = \mathbf{I}$, where $\mathbf{A}$ and $\mathbf{B}$ are the matrices of Problem 5.5.

5.7. Show that matrices of the type

$$\begin{Bmatrix} \cos u & \sin u & 0 & 0 \\ -\sin u & \cos u & 0 & 0 \\ 0 & 0 & \cos v & -\sin v \\ 0 & 0 & \sin v & \cos v \end{Bmatrix}$$

are orthonormal.

5.8. Give another proof of the fact that the inverse of an orthonormal matrix is itself orthonormal by making use of the identity $\mathbf{A}^{-1} = \mathbf{A}_t$.

5.9. Prove that the matrix $\mathbf{A} = (\mathbf{I} - \mathbf{B})(\mathbf{I} + \mathbf{B})^{-1}$, where $\mathbf{B}$ is skew-symmetric, is orthonormal. *Hint.* Use the fact that $\mathbf{I} - \mathbf{B}$ and $\mathbf{I} + \mathbf{B}$ commute.

5.10. Prove that, if $\mathbf{B}$ is skew-symmetric, then $\det(\mathbf{I} - \mathbf{B}) = \det(\mathbf{I} + \mathbf{B})$.

5.11. Invert the following matrices by suitable partitioning—

$$\mathbf{A} = \begin{Bmatrix} 1 & 1 & 1 \\ 2 & 0 & 2 \\ 2 & -2 & 1 \end{Bmatrix}$$

$$\mathbf{B} = \begin{Bmatrix} 2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 2 & -1 \end{Bmatrix}$$

5.12. In Section 5.5 a non-singular matrix

$$\mathbf{Z} = \begin{Bmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 \\ \mathbf{Z}_3 & \mathbf{Z}_4 \end{Bmatrix}$$

was inverted by partitioning. It was tacitly assumed that the square submatrices $\mathbf{Z}_1$ and $\mathbf{Z}_4$ were both invertible. Discuss whether it can be inferred from $\det(\mathbf{Z}) \neq 0$ that $\det(\mathbf{Z}_1) \neq 0$.

5.13. Test the method of partitioning the matrices of a matrix product by subdividing the following matrices according to the rules outlined in Section 5.4, and compare the result with that obtained by ordinary matrix multiplication.

$$\begin{Bmatrix} 1 & -1 & 0 \\ 2 & 1 & 1 \\ 0 & 2 & 1 \\ 3 & -1 & 1 \end{Bmatrix} \begin{Bmatrix} 2 & -1 \\ 1 & 0 \\ 0 & 1 \end{Bmatrix}$$

CHAPTER 6

Linear Transformations and Linear Vector Functions

6.1. Coordinate Transformations

An n-dimensional vector V can be written as a unique linear combination of n n-dimensional linearly independent vectors A_i.

$$V = v_1 A_1 + v_2 A_2 + \ldots + v_n A_n \qquad (6.1)$$

where the A_i form a base for the n-dimensional manifold containing V. The numbers $v_1, \ldots, v_n$ are the coordinates of V in the coordinate system formed by the A_i.

$$V = \{A_1 \ldots A_n\} \begin{Bmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{Bmatrix} \qquad (6.2)$$

Let us now express each A-vector as a linear combination of n new independent vectors $A_1', \ldots, A_n'$. The relationship between the two sets of vectors can be stated in matrix notation as

$$(A_1 \ldots A_n) = (A_1' \ldots A_n')\mathbf{M} \qquad (6.3)$$

where $\mathbf{M}$ is a non-singular n-by-n matrix.

The question then arises, How is the vector V expressed in this new frame of reference? Combining eqns. (6.2) and (6.3), we get

$$V = (A_1 \ldots A_n) \begin{Bmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{Bmatrix} = (A_1' \ldots A_n')\mathbf{M} \begin{Bmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{Bmatrix}$$

$$= (A_1' \ldots A_n') \begin{Bmatrix} v_n' \\ \cdot \\ \cdot \\ \cdot \\ v_n' \end{Bmatrix} \qquad (6.4)$$

in which $v_1', \ldots, v_n'$ are the coordinates of V in the new frame.

By comparing the third and fourth members of eqn. (6.4), and bearing in mind that the A'-vectors are independent, we see that the coordinates of V transform according to the expression

$$\mathbf{M}\left\{\begin{matrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{matrix}\right\} = \left\{\begin{matrix} v_1' \\ \cdot \\ \cdot \\ \cdot \\ v_n' \end{matrix}\right\}$$

After pre-multiplication by $\mathbf{M}^{-1}$ and subsequent transposition (in order to make the transformation formula functionally identical with that of eqn. (6.3)), we arrive at

$$(v_1 \ldots v_n) = (v_1' \ldots v_n')(\mathbf{M}^{-1})_t \qquad . \quad . \quad (6.5)$$

We note that the two expressions are identical in functional form, but that the place of $\mathbf{M}$ in eqn. (6.3) is occupied by $\mathbf{M}_t^{-1}$ in eqn. (6.5). Transformations that behave in this way are said to be *contragredient*. The transformation of the base is conventionally termed *covariant*; that of the coordinates of a vector therefore becomes *contravariant*.

When (and only when) $\mathbf{M}$ is orthonormal, $\mathbf{M}_t^{-1} = \mathbf{M}$, and the distinction between co- and contravariance vanishes.

Note. Contragredient quantities are frequently encountered in the physical sciences. The measure of a distance, for example, is a number indicating how many times a standard unit of length is contained in the given distance. The unit of length and the measure of a distance are contragredient. If we change from centimetres to inches, which are 2·54 times larger, the measure of a distance becomes 2·54 times smaller. On the other hand, the measure of a velocity and the unit of time transform in exactly the same way; they are cogredient.

An excellent example of co- and contravariance is to be found in the method of referring electrical quantities from the primary to the secondary side of an ideal transformer. If the voltage is stepped up in a certain ratio the current will be stepped down in the same ratio (while the power remains invariant).

Exercise 1. From the theory of 3-dimensional vector analysis it is well known that the cosine of the angle between two unit vectors is equal to their scalar product when the coordinate system is Cartesian (i.e. the base vectors are normalized and orthogonal).

Generalizing this method to n dimensions, we may pose the question, How does the scalar product of two vectors transform when the base is subjected to the transformation given by eqn. (6.3)?

Let V_1 and V_2 be two unit vectors. Then

$$\cos \underline{/(V_1, V_2)} = V_{1t}V_2 = V_{1t}\mathbf{I}V_2 = V_{1t}'\mathbf{M}_t^{-1}\mathbf{M}^{-1}V_2' = V_{1t}'\mathbf{G}'V_2' \quad . \ (6.6)$$

where

$$\mathbf{G}' = \mathbf{M}_t^{-1}\mathbf{I}\mathbf{M}^{-1} \qquad . \quad . \quad . \quad . \quad . \ (6.7)$$

is termed the *metric* of the new coordinate system.

$\mathbf{G}'$ is symmetric since

$$\mathbf{G}_t' = (\mathbf{M}_t^{-1}\mathbf{I}\mathbf{M}^{-1})_t = \mathbf{M}_t^{-1}\mathbf{I}_t\mathbf{M}_{tt}^{-1} = \mathbf{M}_t^{-1}\mathbf{I}\mathbf{M}^{-1} = \mathbf{G}' \qquad . \ (6.8)$$

Apparently the unit matrix is the metric of Euclidean space referred to a Cartesian coordinate system. Rearranging eqn. (6.7) to conform to the pattern of eqn. (6.3), we find that

$$\mathbf{I} = \mathbf{M}_t\mathbf{G}'\mathbf{M} \qquad . \quad . \quad . \quad . \quad . \ (6.9)$$

which indicates that the metric is doubly covariant.

Also, from eqn. (6.7) it is readily seen that the metric is invariant ($= \mathbf{I}$) to orthonormal transformations. This is not surprising when we recall that an orthonormal transformation corresponds to replacing one Cartesian system of reference axes by another.

Exercise 2. Let two sets of independent vectors A_1, A_2, A_3 and A_1', A_2', A_3' be given in Euclidean 3-space, and let the A-vectors be orthonormal. The two sets are related by the expression

$$(A_1A_2A_3) = (A_1'A_2'A_3')\mathbf{M} = (A_1'A_2'A_3')\left\{\begin{matrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ -2 & 1 & 0 \end{matrix}\right\}$$

The coordinates of vectors referred to the two systems transform by means of the matrix

$$\mathbf{M}_t^{-1} = \left\{\begin{matrix} 1 & 2 & -1 \\ -1 & -2 & 2 \\ -1 & -1 & 1 \end{matrix}\right\}$$

as given in eqn. (6.5).

The metric of the A-system is $\mathbf{I}$ and that of the A'-system is

$$\mathbf{G}' = \mathbf{M}_t^{-1}\mathbf{M}^{-1} = \left\{\begin{matrix} 6 & -7 & -4 \\ -7 & 9 & 5 \\ -4 & 5 & 3 \end{matrix}\right\}$$

Unit vectors along the first two axes of the A-system will have coordinates (1, 0, 0) and (0, 1, 0), respectively, and transform into (0, 1, −2) and (1, 0, 1), respectively, in the A'-system.

Check by means of eqn. (6.6) that the vectors expressed in the new coordinates are still at right angles to one another and still of unit length.

6.2. Linear Vector Functions

If a correspondence exists between two sets of vectors V and W such that for every vector V there exists a unique vector W, then W is said to be a (vector) function of the (vector) variable V.

$$W = f(V) \qquad . \quad . \quad . \quad . \ (6.10)$$

When, in addition, the condition

$$f(xV_1 + yV_2) = xf(V_1) + yf(V_2) \qquad . \quad . \ (6.11)$$

is satisfied for all values of x and y, the function is termed *linear*.

Note. The relationship defined by eqn. (6.11) is also called a *linear mapping* of V-space *onto* (or *into*) W-space. An alternative name is *homogeneous affinity*, where the word "homogeneous" refers to the fact that a nullvector is mapped into a nullvector. W is termed the *image* of V, and V the *pre-image* of W. Note that V and W need not be of equal dimensions.

Because matrix multiplication is distributive with respect to matrix addition, we see that

$$\mathrm{A}(xV_1 + yV_2) = x\mathrm{A}V_1 + y\mathrm{A}V_2 \quad . \quad . \ (6.12)$$

and we can now formulate a theorem.

Theorem 24. An equation of the form

$$W = \mathrm{A}V \quad . \quad . \quad . \quad . \ (6.13)$$

represents a linear vector function.

Theorem 25. Conversely, any linear vector function can be written as a matrix equation of the type given in Theorem 24.

Proof. Let $V_1, \ldots, V_n$ span n-dimensional vector space and let the V-vectors transform as follows: V_i into W_i.

Further, suppose A to be the required functional matrix. Then the above n relationship may be written in one equation as

$$(W_1W_2 \ldots W_n) = \mathrm{A}(V_1V_2 \ldots V_n) \quad . \quad . \ (6.14)$$

Since the vectors V_i are linearly independent, this equation has a unique solution,

$$\mathrm{A} = (W_1W_2 \ldots W_n)(V_1V_2 \ldots V_n)^{-1} \quad . \quad . \ (6.15)$$

which proves the theorem.

It is significant that the proof does not require the W_i to be independent, nor need A be square.

Exercise. Determine the functional matrix A which transforms (1, 1) into (3, 2), and (0, 1) into (1, 1).

Employing eqn. (6.14), we have

$$\begin{Bmatrix} 3 & 1 \\ 2 & 1 \end{Bmatrix} = \mathrm{A}\begin{Bmatrix} 1 & 0 \\ 1 & 1 \end{Bmatrix}$$

Post multiplication by

$$\begin{Bmatrix} 1 & 0 \\ 1 & 1 \end{Bmatrix}^{-1} = \begin{Bmatrix} 1 & 0 \\ -1 & 1 \end{Bmatrix}$$

yields

$$\mathrm{A} = \begin{Bmatrix} 2 & 1 \\ 1 & 1 \end{Bmatrix}$$

Note. There are two distinct ways of interpreting the linear relationship given by $W = \mathrm{A}V$.

According to the *active* interpretation the vectors W and V are expressed in terms of the same coordinate system, and the application of the transformation to V results in another vector W.

The *passive* point of view considers V and W to be one and the same vector referred to two different coordinate systems (*see* Section 6.1).

6.3. Rank of a Linear Vector Function

Suppose that a linear vector function is given by

$$W = \mathrm{A}V \quad . \quad . \quad . \quad . \ (6.16)$$

The rank of A is called the *rank of the function*. The n unit vectors in n-dimensional V-space transform into the n columns vectors of A. Of these column vectors r are independent.

When A is square and det (A) $\neq 0$, there exists an inverse function A^{-1}, and there is a one-to-one correspondence between the vectors in V-space and those in W-space.

With the concept of linear vector function at our disposal, we can now find another geometric interpretation of the general solution of linear equations written in matrix form (*see* Section 4.2, p. 35).

The algebraic problem is

$$\mathrm{A}X = B \quad . \quad . \quad . \quad . \ (6.17)$$

where the vectors X and B need not be like. Geometrically the question can be formulated, Does a vector (or vectors) exist which, when applied to the linear vector function represented by matrix A, transforms into B?

As X sweeps through all n-space, $\mathrm{A}X$ will map out a space $S(\mathrm{A}X)$ the dimension of which is equal to the rank of A. Unless B belongs to $S(\mathrm{A}X)$ there will not be a corresponding vector X, and the equation will have no solution.

The different possibilities are set out in Table 6.1.

Table 6.1

THE GENERAL SOLUTION OF $\mathrm{A}X = B$

Dimension of X	B does not belong to $S(\mathrm{A}X)$	B belongs to $S(\mathrm{A}X)$	
		$r(\mathrm{A}) = n$	$r(\mathrm{A}) < n$
n	no solution	one unique solution	$X = X_0 +$ nullspace of A

If A is a square matrix the matrices $\frac{1}{2}(\mathrm{A} + \mathrm{A}_t)$ and $\frac{1}{2}(\mathrm{A} - \mathrm{A}_t)$ are respectively symmetric and anti-symmetric. The identity

$$\mathrm{A} \equiv \tfrac{1}{2}(\mathrm{A} + \mathrm{A}_t) + \tfrac{1}{2}(\mathrm{A} - \mathrm{A}_t)$$

shows that a square matrix can always be expressed as the sum of a symmetric and an anti-symmetric (skew) matrix.

An analogous statement can be made concerning linear vector functions that have square functional matrices.

Exercise. Discuss the vector function with the functional matrix

$$\mathbf{A} = \begin{Bmatrix} 1 & 2 \\ 2 & 4 \end{Bmatrix}$$

We note immediately that the columns of A are proportional, hence $r(\mathbf{A}) = 1$. Applying an arbitrary vector (x, y) to A we find that

$$W = \begin{Bmatrix} 1 & 2 \\ 2 & 4 \end{Bmatrix} \begin{Bmatrix} x \\ y \end{Bmatrix} = \begin{Bmatrix} x + 2y \\ 2x + 4y \end{Bmatrix} = (x + 2y) \begin{Bmatrix} 1 \\ 2 \end{Bmatrix}$$

Apparently $S(\mathbf{A}X)$ is a straight line through the origin in the direction determined by the vector (1, 2), and all values of x and y for which $x + 2y$ is a constant will correspond to the same vector.

To find the vector which transforms into (3, 6), we must solve the equation

$$\begin{Bmatrix} 1 & 2 \\ 2 & 4 \end{Bmatrix} \begin{Bmatrix} x \\ y \end{Bmatrix} = \begin{Bmatrix} 3 \\ 6 \end{Bmatrix}$$

By the methods outlined in Section 4.2, p. 35, we get

$$\begin{Bmatrix} x \\ y \end{Bmatrix} = \begin{Bmatrix} 3 \\ 0 \end{Bmatrix} + t \begin{Bmatrix} -2 \\ 1 \end{Bmatrix}$$

where the last term is the nullspace of A.

Fig. 6.1 shows the nullspace of A and also the space $S(\mathbf{A}X)$.

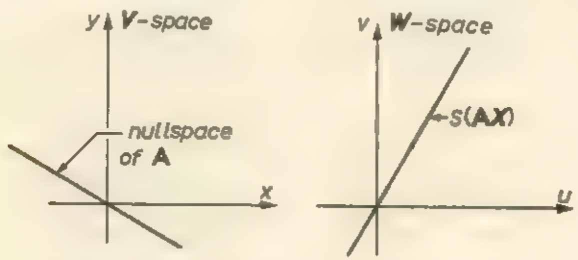


FIG. 6.1. NULLSPACE OF A AND SPACE $S(\mathbf{A}X)$

6.4. Transformation of Functional Matrix

Let
$$W = \mathbf{A}V \qquad (6.18)$$

be a linear vector function, and let the coordinates of W and V transform as described in Section 6.1, namely

$$V' = \mathbf{M}V \quad \text{or} \quad V = \mathbf{M}^{-1}V'$$

and
$$W' = \mathbf{M}W \quad \text{or} \quad W = \mathbf{M}^{-1}W'$$

Substituting these expression in eqn. (6.18) and pre-multiplying by **M** leads to

$$W' = \mathbf{MAM}^{-1}V' = \mathbf{A}'V'$$

This enables us to formulate the next theorem.

Theorem 26. When the vector variables of a linear vector function transform as follows: $V' = \mathbf{M}V$ and $W' = \mathbf{M}W$, the functional matrix A will transform according to the formula

$$\mathbf{A}' = \mathbf{MAM}^{-1} \qquad (6.19)$$

The functional matrix is thus partly covariant and partly contravariant.

Exercise. Let us transpose the members of eqn. (6.19) in order to decide under what conditions symmetry (anti-symmetry) of the functional matrix is an invariant property under the transformation M.

We get

$$\mathbf{A}'_t = \mathbf{M}_t^{-1}\mathbf{A}_t\mathbf{M}_t = \mathbf{M}_t^{-1}\mathbf{A}\mathbf{M}_t \qquad (6.20)$$

when A is symmetric, from which it can be seen that the property of symmetry is invariant when the transformation is orthonormal. The same condition applies to the invariance of skewness, as can readily be inferred from the equation.

It is interesting to note how often orthonormal matrices turn out to be in a class by themselves.

6.5. Definition of a Group—Transformation Groups

Let a finite or infinite set of elements $A, B, C, \ldots$ be given, together with a method for combining them. The combination of two elements is called their *product* and is indicated by simple juxtaposition.

The set will constitute a *group* with respect to the product operation if the following four conditions (the group axioms) are met—

(i) If A and B belong to the set, then AB will also belong to it. This is the condition of *closure*.

(ii) The product of three or more elements is associative.

(iii) The set contains a unit element I such that

$$AI = A(= IA)$$

where A is any element of the set.

(iv) For every element A of the set there exists an inverse element A^{-1} belonging to the set such that

$$AA^{-1} = I(= A^{-1}A)$$

If, in addition, $AB = BA$, where A and B are any elements, the group is said to be *commutative* (or Abelian).

A group in which all the elements can be derived from one of them by repetition of the product operation, is termed *cyclic*. Cyclic groups are always Abelian.

If all the elements of a group H are also elements of a group G, then H is a *subgroup* of G.

In this book we shall only consider groups whose elements are matrices with matrix multiplication as the law of combination. Therefore, since matrix multiplication is generally associative, we need not investigate condition (ii) when verifying that a set of matrices forms a group.

Two groups G (with elements $A, B, C, \ldots$) and G' (with elements $A', B', C', \ldots$) are *isomorphic* if a one-to-one correspondence can be found between their elements $A \leftrightarrow A'$, $B \leftrightarrow B', \ldots$ such that the identities $AB = C$ and $A'B' = C'$ imply one another. Isomorphic groups are structurally identical.

Exercise 1. From Section 5.3, p. 53, it is clear that all n-by-n orthonormal matrices form an infinite group with respect to matrix multiplication: the product of two orthonormal matrices is orthonormal, the unit matrix is orthonormal and every orthonormal matrix has an inverse (its transpose) which also is orthonormal.

Exercise 2. An example of a finite group comprising four elements is the rotation group

$$\mathbf{I} = \begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}, \quad -\mathbf{I} = \begin{Bmatrix} -1 & 0 \\ 0 & -1 \end{Bmatrix}, \quad \mathbf{J} = \begin{Bmatrix} 0 & -1 \\ 1 & 0 \end{Bmatrix}, \quad -\mathbf{J} = \begin{Bmatrix} 0 & 1 \\ -1 & 0 \end{Bmatrix}$$

This group is cyclic since all its elements can be written as powers of $\mathbf{J}$: $\mathbf{J}^1 = \mathbf{J}$, $\mathbf{J}^2 = -\mathbf{I}$, $\mathbf{J}^3 = -\mathbf{J}$, and $\mathbf{J}^4 = \mathbf{I}$.

The rotation group is isomorphic with the group consisting of the numbers $+1, j, -1$, and $-j$ with ordinary multiplication of complex numbers as group operation.

Exercise 3. The following sets of matrices form groups—

(i) All n-by-n matrices with determinants $\neq 0$.
(ii) All n-by-n matrices with determinants $= \pm 1$.
(iii) All n-by-n matrices with determinants $= +1$.

Set (iii) is a subgroup of set (ii), which in turn is a subgroup of set (i). The group of orthonormal matrices is a subgroup of set (ii).

The set of all n-by-n matrices with determinants $= -1$ does not form a group; for one thing, the set does not include the identity matrix, and secondly, the condition of closure is not satisfied (compare Theorem 15, p. 26).

An extremely useful property of the group concept is that it enables us to classify transformations. As we shall see in Chapter 12, groups of transformations and invariance with respect to such groups play an essential part in the definition of a tensor.

In order to relate the two ideas, transformation and group, we shall have to decide what the elements of the group are to be and how to define the group operation.

A linear transformation is uniquely determined by its matrix (see Theorems 24 and 25, p. 62). It would therefore seem natural to define the elements of the transformation group to be the transformation matrices.

As the group operation we shall choose the consecutive application of two transformations. It is easy to prove that this operation is performed algebraically by multiplying the corresponding transformation matrices. Let us apply a linear transformation to the axes of a coordinate system in such a way that the coordinates of a vector V transform according to the formula given in Section 6.1—

$$V' = \mathbf{M}V \qquad (6.21)$$

Further, let us apply another transformation to the new axes, thus taking V' into V'' as follows—

$$V'' = \mathbf{N}V' \qquad (6.22)$$

Substituting eqn. (6.21) in eqn. (6.22), we find that

$$V'' = \mathbf{NM}V \qquad (6.23)$$

Thus the composition of the two transformations $\mathbf{M}$ and $\mathbf{N}$ (i.e. the transformation taking V directly into V'') is represented by the transformation matrix $\mathbf{NM}$.

Having now chosen the group elements and decided upon a group operation, we still have to demonstrate that the four group axioms are satisfied.

The condition of closure is obviously complied with: the combination of two linear transformations is a linear transformation whose matrix is the product of those of the component transformations.

Associativity is guaranteed by virtue of the general associativity of matrix multiplication.

The unit element of a transformation group (corresponding to the act of "doing nothing") is the identity transformation in which each coordinate axis transforms into itself. It is represented by the unit matrix.

Finally, to comprise a group, each element must possess a unique inverse. Hence, only one-to-one transformations, which are always represented by square matrices with non-zero determinants, can be the elements of a transformation group.

6.6. Eigenvectors and Eigenvalues

When a vector function $\mathbf{A}$ is applied to a vector V, the result is a vector $\mathbf{A}V$ which generally differs from V both in magnitude and

direction. The question naturally arises, Are there vectors which, when operated upon by the functional matrix **A**, do not change their direction?

Algebraically stated,

$$\mathbf{A}X = \lambda X \quad . \quad . \quad . \quad . \quad (6.24)$$

A solution to eqn. (6.24) is called an *eigenvector* (characteristic vector) of the linear vector function **A**; and the corresponding λ, an *eigenvalue* (latent root, characteristic value). From the point of view of an eigenvector, the vector function is equivalent to a simple scalar multiplier λ.

We shall see later that eigenvectors are fundamentally important in many fields of algebra, geometry, applied mathematics and mathematical engineering. Problems of stability in complicated interconnected systems, for instance, can often be reduced to eigenvalue problems.

To solve eqn. (6.24) we rearrange it as follows—

$$(\mathbf{A} - \lambda\mathbf{I})X = 0 \quad . \quad . \quad . \quad (6.25)$$

From the discussion of nullspaces in Section 4.3, p. 39, we know that this equation has a non-trivial solution only when its determinant vanishes.

In expanded form the determinant $\det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{R}(\lambda))$ is an nth-degree polynomial in λ (the *reduction polynomial*) which may have up to n real roots.

Assuming λ_1 to be a real root of the reduction polynomial (i.e. an eigenvalue of **A**), we determine the associated eigenvector from eqn. (6.25) by substituting λ_1 for λ.

If V_1 is an eigenvector of a given vector function **A**, then kV_1 is also an eigenvector of the function because

$$\mathbf{A}(kV_1) = k\mathbf{A}V_1 = k\lambda_1 V_1 = \lambda_1(kV_1) \quad . \quad . \quad (6.26)$$

This follows from the linearity of the function.

Let V_1 and V_2 be two non-collinear eigenvectors with the same eigenvalue λ. Then

$$\mathbf{A}(k_1V_1 + k_2V_2) = k_1\lambda V_1 + k_2\lambda V_2 = \lambda(k_1V_1 + k_2V_2) \quad (6.27)$$

Hence, we have proved the next theorem.

Theorem 27. Any linear combination of two or more eigenvectors with the same eigenvalue is itself an eigenvector.

In view of this theorem it would be more logical to speak of an *eigendirection* or an *eigenspace* in connexion with a given eigenvalue. The rank of such an eigenspace is equal to the multiplicity of the

associated eigenvalue (and also equal to the nullity of the corresponding reduction matrix $\mathbf{R}(\lambda)$).

Suppose V_1 and V_2 are two eigenvectors of a linear vector function with symmetrical functional matrix, and let λ_1 and λ_2 be the corresponding non-equal eigenvalues.

We now have

$$\mathbf{A}V_1 = \lambda_1 V_1 \quad . \quad . \quad . \quad . \quad (6.28)$$

and

$$\mathbf{A}V_2 = \lambda_2 V_2 \quad . \quad . \quad . \quad . \quad (6.29)$$

Pre-multiplying these two equations by V_{2t} and V_{1t} respectively, we find that

$$V_{2t}\mathbf{A}V_1 = \lambda_1 V_{2t}V_1 \quad . \quad . \quad . \quad (6.30)$$

and

$$V_{1t}\mathbf{A}V_2 = \lambda_2 V_{1t}V_2 \quad . \quad . \quad . \quad (6.31)$$

By transposing both members of eqn. (6.30), we have

$$V_{1t}\mathbf{A}_tV_2 = V_{1t}\mathbf{A}V_2 = \lambda_1 V_{1t}V_2 \quad . \quad . \quad (6.32)$$

This process does not alter the equations, which are merely scalar identities. Therefore, since the left-hand members of eqns. (6.31) and (6.32) are identical, the right-hand members must be so too. Thus

$$\lambda_1 V_{1t}V_2 = \lambda_2 V_{1t}V_2 \quad . \quad . \quad . \quad (6.33)$$

and because $\lambda_1 \neq \lambda_2$, we must needs conclude that $V_{1t}V_2 = 0$.

In Cartesian coordinates, therefore, we can formulate Theorem 28.

Theorem 28. Eigenvectors of a symmetric linear vector function corresponding to different eigenvalues are orthogonal (i.e. their scalar product is zero).

Exercise 1. To find the eigenvalues and eigenvectors of the function

$$\mathbf{A} = \begin{Bmatrix} 2 & -2 & 1 \\ -1 & 3 & -1 \\ 2 & -4 & 3 \end{Bmatrix}$$

we look for the roots in the reduction polynomial

$$|\mathbf{A} - \lambda\mathbf{I}| = |\mathbf{R}(\lambda)| = \begin{vmatrix} 2-\lambda & -2 & 1 \\ -1 & 3-\lambda & -1 \\ 2 & -4 & 3-\lambda \end{vmatrix}$$
$$= -\lambda^3 + 8\lambda^2 - 13\lambda + 6 = 0$$

They are $\lambda_1 = 6$ and $\lambda_2 = \lambda_3 = 1$.

To determine the corresponding eigenvectors we must solve eqn. (6.25): $(\mathbf{A} - \lambda_1\mathbf{I})X = 0$. For $\lambda = 6$ we have

$$\begin{Bmatrix} -4 & -2 & 1 \\ -1 & -3 & -1 \\ 2 & -4 & -3 \end{Bmatrix} \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix} \quad . \quad . \quad . \quad (6.34)$$

$\lambda = 6$ is a single root; hence the nullity of the reduction matrix is 1 and its rank is 2 ($= 3 - 1$). The solutions of eqn. (6.34) will occupy a 1-space (the eigendirection). Hence, by means of the technique developed in Section 4.3, we get

$$V_1 = t\,.\begin{Bmatrix} 1 \\ -1 \\ 2 \end{Bmatrix}$$

As a check we compute

$$AV_1 = \begin{Bmatrix} 2 & -2 & 1 \\ -1 & 3 & -1 \\ 2 & -4 & 3 \end{Bmatrix}\begin{Bmatrix} t \\ -t \\ 2t \end{Bmatrix} = \begin{Bmatrix} 6t \\ -6t \\ 12t \end{Bmatrix}$$

which proves our calculations to be correct.

In the case of the double root $\lambda = 1$, we must solve the equation

$$\begin{Bmatrix} 1 & -2 & 1 \\ -1 & 2 & -1 \\ 2 & -4 & 2 \end{Bmatrix}\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix}$$

We now have $r(\mathrm{A} - \mathrm{I}) = 1$ (which is clear from the fact that all the columns of $\mathbf{A} - \mathbf{I}$ are proportional). The eigenspace is 2-dimensional and determined by the equation

$$x - 2y + z = 0$$

Any vector from this space is an eigenvector with the eigenvalue 1. We select (arbitrarily) the two orthogonal (and therefore independent) vectors

$$V_2 = \begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix} \quad \text{and} \quad V_3 = \begin{Bmatrix} -1 \\ 0 \\ 1 \end{Bmatrix}$$

The reader should test that any linear combination of V_2 and V_3 is an eigenvector with eigenvalue 1.

Exercise 2. Determine the eigenvalues and eigenvectors of the symmetric vector function

$$\begin{Bmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{Bmatrix}$$

The reduction polynomial is

$$|\mathrm{R}(\lambda)| = -\lambda^3 + 18\lambda^2 - 99\lambda + 162 = 0$$

which has three distinct roots, 3, 6, and 9.

The corresponding eigenvectors are determined in the manner described in Exercise 1, and are found to be

$$V_1 = \begin{Bmatrix} 1 \\ 2 \\ 2 \end{Bmatrix}, \quad V_2 = \begin{Bmatrix} 2 \\ 1 \\ -2 \end{Bmatrix} \quad \text{and} \quad V_3 = \begin{Bmatrix} 2 \\ -2 \\ 1 \end{Bmatrix}$$

In accordance with Theorem 28, the three eigenvectors are mutually orthogonal.

Exercise 3. A very special example of a vector function is the *indentity function* represented by a unit matrix. In this case the reduction polynomial is simply

$$(1 - \lambda)^n = 0$$

which has an n-fold root $\lambda = 1$. The reduction matrix $\mathbf{A} - \lambda\mathbf{I}$ degenerates into an n-by-n nullmatrix (the nullity of which is n), and any vector is an eigenvector.

Exercise 4. Discuss the 3-dimensional skew (anti-symmetric) vector function

$$\mathrm{A} = \begin{Bmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{Bmatrix}$$

The rank of A is 2 and the reduction polynomial is

$$\det(\mathrm{A} - \lambda\mathrm{I}) = \begin{Bmatrix} -\lambda & c & -b \\ -c & -\lambda & a \\ b & -a & -\lambda \end{Bmatrix} = -\lambda(\lambda^2 + a^2 + b^2 + c^2) = 0$$

The function has one real eigenvalue $\lambda = 0$ (this is obvious from the fact that $\det(\mathrm{A}) = 0$), and the corresponding eigenvector is found by solving the equation

$$\begin{Bmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{Bmatrix}\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} 0 + cy - bz \\ -cx + 0 + az \\ bx - ay + 0 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix} \quad . \tag{6.35}$$

which yields

$$V = t\,.\begin{Bmatrix} a \\ b \\ c \end{Bmatrix}$$

We note that $V_t\mathrm{A} = 0_t$ (this follows because V is an eigenvector with zero eigenvalue, and A is skew). Hence, V and $\mathbf{A}X$, where X is any vector, will be orthogonal—

$$V_t(\mathrm{A}X) = (V_t\mathrm{A})X = 0_tX = 0$$

Also, X and $\mathbf{A}X$ are always orthogonal, since $X_t\mathbf{A}X$ vanishes identically (we shall encounter this property of skew matrices in connexion with quadratic forms in Section 7.1, where a proof is given).

The reader who is familiar with 3-dimensional vector algebra will recognize the middle member of eqn. (6.35) as the expression for the cross-product of the vector X and a vector with coordinates (a, b, c).

As a further exercise the reader might investigate how many of the above-mentioned properties of skew vector functions in 3-space carry over into the n-dimensional case.

Exercise 5. In this exercise we shall discuss vector functions with orthonormal matrices in Cartesian coordinate systems. Bearing in mind our previous experience with matrices of this type, it would seem reasonable to expect that such functions should exhibit rather special characteristics; this turns out to be the case.

(i) The magnitude of a vector is invariant to the application of an orthonormal vector function—

$$(\mathrm{A}V)_t(\mathrm{A}V) = V_t\mathrm{A}_t\mathrm{A}V = V_t\mathrm{I}V = V_tV$$

i.e. the scalar product of a vector by itself is invariant.

(ii) Similarly the scalar product of any two vectors V_1 and V_2 is also invariant when subjected to an orthonormal transformation.

From these properties we can infer that the angle between two vectors is equal to the angle between their images: the angle between two vectors is also invariant to an orthonormal transformation.

(iii) An n-by-n orthonormal functional matrix (where n is odd) with a determinant equal to $+1$ has an eigenvalue $\lambda = +1$. This is equivalent to proving that

$$\det(\mathbf{A} - \mathbf{I}) = 0 \qquad . \quad . \quad . \quad . \quad (6.36)$$

Making use of eqn. (3.6), p. 24, and Theorem 15, p. 26, we can derive the expression

$$\det(\mathbf{A} - \mathbf{I}) = \det(\mathbf{A}_t)\det(\mathbf{A} - \mathbf{I}) = \det(\mathbf{A}_t(\mathbf{A} - \mathbf{I})) = \det(\mathbf{I} - \mathbf{A}_t)$$
$$= \det(-(\mathbf{A} - \mathbf{I})_t) = (-1)^n \det(\mathbf{A} - \mathbf{I}) = -\det(\mathbf{A} - \mathbf{I})$$

and comparing the first and last terms in this long chain of identities, it is clear that eqn. (6.36) is valid.

By an exactly analogous train of reasoning we can demonstrate that when $\det(\mathbf{A}) = -1$ the functional matrix has the eigenvalue -1.

For the special case $n = 3$ an orthonormal function represents a rigid rotation in space. This immediately explains why the magnitude of a vector and the angle between two vectors remain unaltered. The eigendirection corresponding to the eigenvalue $\lambda = 1$ is the axis of rotation. When the functional determinant is -1, the rotation is coupled with a reflexion.

It is instructive to note how the proof of (iii) breaks down for even dimensions by leading us to the sterile identity of the reduction determinant with itself.

6.7. Invariance of Reduction Polynomial

It has been shown in Section 6.4 that the functional matrix of a linear function transforms according to the formula

$$\mathbf{A}' = \mathbf{MAM}^{-1} \qquad . \quad . \quad . \quad . \quad (6.37)$$

when the vectors transform as follows

$$V' = \mathbf{M}V$$

In the new coordinate system the reduction matrix is

$$\mathbf{R}'(\lambda) = \mathbf{A}' - \lambda\mathbf{I} = \mathbf{MAM}^{-1} - \lambda\mathbf{I} = \mathbf{MAM}^{-1} - \lambda\mathbf{MIM}^{-1}$$
$$= \mathbf{M}(\mathbf{A} - \lambda\mathbf{I})\mathbf{M}^{-1} = \mathbf{MR}(\lambda)\mathbf{M}^{-1} \qquad . \quad . \quad . \quad (6.38)$$

Thus, the reduction matrix transforms in exactly the same way as the functional matrix. By using Theorem 15, p. 26, we find that

$$\det(\mathbf{R}'(\lambda)) = \det\mathbf{M}\det(\mathbf{R}(\lambda))\det(\mathbf{M}^{-1})$$
$$\det\mathbf{M}\det(\mathbf{R}(\lambda))/\det(\mathbf{M}) = \det(\mathbf{R}(\lambda)).$$

Theorem 29. The reduction polynomial $\det(\mathbf{R}(\lambda))$, belonging to a vector function with a square functional matrix, is invariant to any non-singular linear coordinate transformation.

In expanded form the reduction polynomial will be

$$\det(\mathbf{R}(\lambda)) = (-1)^n\lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0$$

where the c's are scalar constants and n is the order of the determinant. Since the polynomial is invariant to all invertible linear transformations for all values of λ, each of the constants c must be invariant.

Two of these scalar invariants are of particular interest—

(i) c_0, the constant term of the polynomial, is equal to $\det(\mathbf{A})$. When $c_0 = 0$, the function is singular and $\lambda = 0$ is an eigenvalue. All the corresponding eigenvectors fill the nullspace of $\mathbf{A}$.

If the nullity of $\mathbf{A}$ is $p \geqslant 1$, the constants up to and including c_{p-1} vanish, and $\lambda = 0$ is a p-fold solution of the polynomial.

(ii) c_{n-1}, the coefficients of λ^{n-1}, represents the sum of the elements along the main diagonal of the functional matrix multiplied by the factor $(-1)^{n-1}$:

$$c_{n-1} = (-1)^{n-1}(a_{11} + a_{22} + \ldots + a_{nn})$$

Some authors call it the *trace* of the matrix $\mathbf{A}$; others use the German word *Spur*. The process of forming the trace of a square matrix is called *contraction*.

Exercise. Take the function discussed in Exercise 2, Section 6.6, whose matrix was

$$\mathbf{A} = \begin{Bmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{Bmatrix}$$

and apply transformation

$$\mathbf{M} = \begin{Bmatrix} 0 & 1 & -2 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{Bmatrix}$$

to the coordinates. The inverse of $\mathbf{M}$ is readily calculated to be

$$\mathbf{M}^{-1} = \begin{Bmatrix} 1 & 2 & -1 \\ -1 & -2 & 2 \\ -1 & -1 & 1 \end{Bmatrix}$$

and the transformed functional matrix is

$$\mathbf{A}' = \mathbf{MAM}^{-1} = \begin{Bmatrix} 0 & -12 & 10 \\ 6 & 17 & -10 \\ 3 & 4 & 1 \end{Bmatrix}$$

We can now compute the reduction polynomial

$$\det(\mathbf{R}'(\lambda)) = -\lambda^3 + 18\lambda^2 - 99\lambda + 162$$

which is identical with the expression derived in Exercise 2, Section 6.6.

Note that the trace of both $\mathbf{A}$ and $\mathbf{A}'$ is $+18$, and that $\det(\mathbf{A}) = \det(\mathbf{A}') = 162$.

To conclude this section we shall prove a theorem concerning symmetric functional matrices.

Theorem 30. A linear vector function with a symmetric n-by-n functional matrix has n real eigenvalues.

Proof. Let $\mathbf{A}$ be symmetric. If we employ for a moment the field of complex numbers, the reduction polynomial will always have n roots. Suppose now that λ is a complex root of the equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, and hence an eigenvalue of the vector function $\mathbf{A}$. Since the coefficients of the reduction polynomial are real, the conjugate λ_c of λ must also be an eigenvalue of $\mathbf{A}$.

If V is an eigenvector corresponding to λ, then V_c is an eigenvector corresponding to λ_c, and the coordinates of V_c are the conjugates of the corresponding coordinates of V. Thus we have

$$\mathbf{A}V = \lambda V$$

and

$$\mathbf{A}V_c = \lambda_c V_c$$

Pre-multiplying these equations by V_{ct} and V_t respectively and rearranging, we get

$$\lambda = V_{ct}\mathbf{A}V/(V_{ct}V) \quad . \quad . \quad . \quad (6.39)$$

and

$$\lambda_c = V_t\mathbf{A}V_c/(V_tV_c) \quad . \quad . \quad . \quad (6.40)$$

By transposing the right-hand member of the scalar eqn. (6.40) it is seen to be identical with the right-hand member of eqn. (6.39). Thus $\lambda = \lambda_c$, and therefore λ must be real.

6.8. Reduction of Functional Matrix to Diagonal Form

We shall now set ourselves the problem of finding a coordinate transformation $\mathbf{M}$ which will reduce the matrix of a vector function to a diagonal matrix

$$\mathbf{B} = \begin{Bmatrix} b_1 & 0 & . & . & 0 \\ 0 & b_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & b_n \end{Bmatrix}$$

This means that $\mathbf{M}$ must transform $\mathbf{A}$ into $\mathbf{B}$ as follows—

$$\mathbf{MAM}^{-1} = \mathbf{B}$$

Pre-multiplying by $\mathbf{M}^{-1}$, we get

$$\mathbf{AM}^{-1} = \mathbf{M}^{-1}\mathbf{B}$$

The column vector V_i of the matrix $\mathbf{M}^{-1}$ must satisfy the equation

$$\mathbf{A}V_i = \mathbf{M}^{-1}B_i = \mathbf{M}^{-1}\begin{Bmatrix} 0 \\ . \\ . \\ b_i \\ . \\ 0 \end{Bmatrix} = b_iV_i \quad . \quad . \quad (6.41)$$

where B_i is the ith column of matrix $\mathbf{B}$.

Eqn. (6.41) clearly shows (*see* eqn. (6.24)) that the columns of $\mathbf{M}^{-1}$ are the eigenvectors of the functional matrix $\mathbf{A}$, and the diagonal elements of $\mathbf{B}$ are the corresponding eigenvalues.

We can now formulate Theorem 31.

Theorem 31. Any n-by-n functional matrix with n real eigenvalues $\lambda_1, . . ., \lambda_n$ can be reduced to the diagonal form (we shall coin a verb *to diagonalize*).

$$\mathbf{A}' = \begin{Bmatrix} \lambda_1 & 0 & . & . & 0 \\ 0 & \lambda_2 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \lambda_n \end{Bmatrix}$$

by a transformation $\mathbf{M}$, where the column vectors V_i of $\mathbf{M}^{-1}$ are the eigenvectors belonging to the eigenvalues λ_i.

Theorem 32. With the help of Theorem 30 we can state that a symmetric vector function can always be diagonalized.

Note. The reason for calling the expanded determinant

$$\det(\mathbf{R}(\lambda)) = \det(\mathbf{A} - \lambda\mathbf{I})$$

the *reduction polynomial* ought now to be apparent.

Exercise 1. Diagonalize the functional matrix

$$\mathbf{A} = \begin{Bmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{Bmatrix}$$

of Exercise 2, Section 6.6.

The eigenvalues of this matrix are 3, 6 and 9, and the eigenvectors are

$$V_1 = \begin{Bmatrix} 1 \\ 2 \\ 2 \end{Bmatrix}, \quad V_2 = \begin{Bmatrix} 2 \\ 1 \\ -2 \end{Bmatrix} \quad \text{and} \quad V_3 = \begin{Bmatrix} 2 \\ -2 \\ 1 \end{Bmatrix}$$

respectively. Joined together, they form the inverse of the required transformation matrix,

$$\mathbf{M}^{-1} = \begin{Bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{Bmatrix}$$

from which we calculate

$$\mathbf{M} = \frac{1}{9}\begin{Bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{Bmatrix}$$

The observant reader will have noted that $\mathbf{M}$ and $\mathbf{M}^{-1}$ are proportional; this is because $\mathbf{M}^{-1}$ is orthogonal and symmetric, and its column vectors are of equal length.

Applying the transformation to the functional matrix, we find

$$\begin{aligned} \mathrm{A}' &= \mathbf{MAM}^{-1} \\ &= \frac{1}{9}\begin{Bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{Bmatrix}\begin{Bmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{Bmatrix}\begin{Bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{Bmatrix} \\ &= \begin{Bmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 9 \end{Bmatrix} \end{aligned}$$

In practice it is quite unnecessary to compute $\mathbf{MAM}^{-1}$, as we have done here, since the diagonalized form matrix is obtained as soon as the eigenvalues have been determined. In many cases there is no need even to calculate $\mathbf{M}$ and $\mathbf{M}^{-1}$.

Exercise 2. In Exercise 1, Section, 6.6, the functional matrix

$$\mathrm{A} = \begin{Bmatrix} 2 & -2 & 1 \\ -1 & 3 & -1 \\ 2 & -4 & 3 \end{Bmatrix}$$

was shown to have the eigenvalues 6, 1 and 1. The single root yielded the eigenvector

$$V_1 = \begin{Bmatrix} 1 \\ -1 \\ 2 \end{Bmatrix}$$

and from the 2-dimensional eigenspace corresponding to the double eigenvalue 1, the vectors

$$V_2 = \begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix} \quad \text{and} \quad V_3 = \begin{Bmatrix} -1 \\ 0 \\ 1 \end{Bmatrix}$$

were chosen.

In order to demonstrate that any set of independent eigenvectors selected from the eigenspace of a multiple eigenvalue will perform the diagonalization of A, we shall use the vectors V_2 and $V_3' = V_2 + tV_3$ to build the transformation matrix $\mathbf{M}^{-1}$.

Thus
$$\mathbf{M}^{-1} = \begin{Bmatrix} 1 & 1 & 1-t \\ -1 & 1 & 1 \\ 2 & 1 & 1+t \end{Bmatrix}$$

from which we compute

$$\mathbf{M} = \frac{1}{5t}\cdot\begin{Bmatrix} t & -2t & t \\ 3+t & -1+3t & -2+t \\ -3 & 1 & 2 \end{Bmatrix}$$

The presence of the parameter t in the denominator is a warning that t cannot be allowed to be zero; for when $t = 0$, V_2 and V_3' are no longer independent.

The transformation $\mathbf{M}$ takes A into

$$\mathrm{A}' = \mathbf{MAM}^{-1} = \begin{Bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Bmatrix}$$

all the t's cancelling out in the process.

Exercise 3. Reduce to diagonal form the matrix

$$\mathrm{A} = \begin{Bmatrix} 0 & 1 \\ 1 & 0 \end{Bmatrix}$$

The reduction polynomial is

$$\det(\mathbf{R}(\lambda)) = \begin{vmatrix} -\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - 1 = 0$$

Hence, $\lambda_1 = +1$ and $\lambda_2 = -1$, and the corresponding eigenvectors are readily found to be

$$V_1 = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} \quad \text{and} \quad V_2 = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}$$

respectively. Thus

$$\mathbf{M}^{-1} = \begin{Bmatrix} 1 & 1 \\ 1 & -1 \end{Bmatrix} \quad \text{and}\ \mathbf{M} = \tfrac{1}{2}\cdot\begin{Bmatrix} 1 & 1 \\ 1 & -1 \end{Bmatrix}$$

and A transforms into

$$\begin{aligned} \mathrm{A}' = \mathbf{MAM}^{-1} &= \tfrac{1}{2}\cdot\begin{Bmatrix} 1 & 1 \\ 1 & -1 \end{Bmatrix}\begin{Bmatrix} 0 & 1 \\ 1 & 0 \end{Bmatrix}\begin{Bmatrix} 1 & 1 \\ 1 & -1 \end{Bmatrix} \\ &= \begin{Bmatrix} 1 & 0 \\ 0 & -1 \end{Bmatrix} \end{aligned}$$

The functional matrix A is symmetric, and the two eigenvalues are not equal. This means (*see* Theorem 28, p. 69) that the column vectors of $\mathbf{M}^{-1}$ are orthogonal, and we can refine our solution by making $\mathbf{M}^{-1}$ orthonormal. The modified transformation matrix is

$$\mathbf{N}^{-1} = \begin{Bmatrix} \tfrac{1}{2}\sqrt{2} & \tfrac{1}{2}\sqrt{2} \\ \tfrac{1}{2}\sqrt{2} & -\tfrac{1}{2}\sqrt{2} \end{Bmatrix}$$

$\mathbf{N}^{-1}$ is also symmetric, and hence

$$\mathbf{N} = (\mathbf{N}^{-1})_t = (\mathbf{N}_t)^{-1} = \mathbf{N}^{-1}$$

Note. Didactically, problems in 2-space are well suited to give the reader a clear understanding of the geometric meaning of his algebraic manipulations. Assuming that the vector function of Exercise 3 is referred to Cartesian coordinates, it is a simple matter to visualize how it operates.

The function A reflects the whole xy-plane about the line $y = x$. This line is therefore an eigendirection with eigenvalue $+1$. Any line perpendicular to $y = x$ remains invariant, except for the fact that it is swung round to point the other way; hence the eigenvalue of the eigendirection $y = -x$ is equal to -1. Applying the orthonormal transformation $\mathbf{N}$ to the function merely rotates the axes to lie along the two orthogonal eigendirections.

6.9. The Cayley–Hamilton Theorem

As a necessary preliminary to the proof of the Cayley–Hamilton theorem, we shall investigate how the powers of a functional matrix are affected by coordinate transformations.

Raising both members of the transformation eqn. (6.19) to the power p, we get

$$\begin{aligned}(\mathbf{A}')^p &= (\mathbf{MAM}^{-1})^p = (\mathbf{MAM}^{-1})(\mathbf{MAM}^{-1})\ldots(\mathbf{MAM}^{-1})\\ &= \mathbf{MAM}^{-1}\mathbf{MAM}^{-1}\ldots\mathbf{MAM}^{-1}\\ &= \mathbf{MAIAI}\ldots\mathbf{IAM}^{-1} = \mathbf{MA}^p\mathbf{M}^{-1} = \mathbf{M}(\mathbf{A}^p)\mathbf{M}^{-1} \quad (6.42)\end{aligned}$$

which shows that $\mathbf{A}^p$ transforms in exactly the same way as $\mathbf{A}$.

Substituting $\mathbf{A}$ for λ and $\mathbf{I}$ ($\stackrel{D}{=} \mathbf{A}^0$) for λ^0 in the reduction polynomial, we obtain a *matrix reduction polynomial* which we shall denote by $\mathbf{R}(\mathbf{A})$, and we can now state the Cayley–Hamilton theorem.

Theorem 33 (*Cayley–Hamilton Theorem*). The matrix reduction polynomial of a functional matrix vanishes identically. This theorem is often paraphrased to read: A functional matrix satisfies its own reduction polynomial.

To prove the theorem, let $\mathbf{A}$ be an n-by-n functional matrix with n real eigenvalues, and let $\mathbf{M}$ be a diagonalizing transformation such that

$$\mathbf{MAM}^{-1} = \mathbf{A}_d$$

where $\mathbf{A}_d$ is a diagonal matrix whose n non-zero elements are the n eigenvalues. Finally, let $\mathbf{R}(\mathbf{A})$ denote the matrix reduction polynomial corresponding to the vector function $\mathbf{A}$.

Applying transformation $\mathbf{M}$ to $\mathbf{R}(\mathbf{A})$, we derive

$$\begin{aligned}\mathbf{MR}(\mathbf{A})\mathbf{M}^{-1} &= \mathbf{M}[(-1)^n\mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + \ldots + c_1\mathbf{A} + c_0\mathbf{I}]\mathbf{M}^{-1}\\ &= (-1)^n\mathbf{MA}^n\mathbf{M}^{-1} + c_{n-1}\mathbf{MA}^{n-1}\mathbf{M}^{-1} + \ldots\\ &\qquad + c_1\mathbf{MAM}^{-1} + c_0\mathbf{MIM}^{-1}\\ &= (-1)^n\mathbf{A}_d^n + c_{n-1}\mathbf{A}_d^{n-1} + \ldots + c_1\mathbf{A}_d + c_0\mathbf{I} \quad . (6.43)\end{aligned}$$

All the off-diagonal elements of this matrix equation are zero, so we need only study the elements along the main diagonal. Recalling that, if $\mathbf{A}_d$ is a diagonal matrix with element a_{di} in position ii, then $\mathbf{A}_d^p$ will also be diagonal with a_{di}^p in position ii, it is clear that the element in row i and column i of the last member of eqn. (6.43) is

$$(-1)^n \cdot \lambda_i^n + c_{n-1}\lambda_i^{n-1} + \ldots + c_1\lambda_i + c_0 \quad . (6.44)$$

and, since λ_i is a root of the reduction polynomial, this expression is equal to zero.

We conclude, therefore, that

$$\mathbf{MR}(\mathbf{A})\mathbf{M}^{-1} = \mathbf{0}$$

from which we derive, by pre-multiplication by $\mathbf{M}^{-1}$ and post-multiplication by $\mathbf{M}$, that

$$\mathbf{R}(\mathbf{A}) = \mathbf{0}$$

thus proving the theorem.

The theorem is still valid when we weaken the condition that $\mathbf{A}$ shall have n real eigenvalues so as to permit the reduction polynomial to have complex roots.

With the aid of complex numbers we can pursue the same train of reasoning to arrive at the same result.

Exercise. Test the Cayley-Hamilton theorem in connexion with the vector function

$$\mathbf{A} = \begin{Bmatrix}0 & -1\\ 1 & 0\end{Bmatrix}$$

The reduction polynomial is det $\mathbf{R}(\lambda) = \lambda^2 + 1$, which has no real roots (the function corresponds to a 90° anti-clockwise rotation of the xy-plane about the origin, an operation which leaves no vector unchanged).

Substituting in the polynomial,

$$\begin{aligned}\mathbf{A}^2 + \mathbf{I} &= \begin{Bmatrix}0 & -1\\ 1 & 0\end{Bmatrix}\begin{Bmatrix}0 & -1\\ 1 & 0\end{Bmatrix} + \begin{Bmatrix}1 & 0\\ 0 & 1\end{Bmatrix}\\ &= \begin{Bmatrix}-1 & 0\\ 0 & -1\end{Bmatrix} + \begin{Bmatrix}1 & 0\\ 0 & 1\end{Bmatrix} = \begin{Bmatrix}0 & 0\\ 0 & 0\end{Bmatrix}\end{aligned}$$

PROBLEMS

6.1. Determine the matrix of a linear vector function which maps (1, 1) into (2, 0, 1), and (0, 1) into (1, −1, 0). Is this linear function one-to-one? Can it be inverted?

6.2. Determine the functional matrices of the following linear vector functions—

$$\left.\begin{aligned}(1, 0, 0) &\text{ into } (0, -1, 1)\\ (1, 0, 1) &\text{ into } (1, 0, 0)\\ (0, 1, -1) &\text{ into } (0, 1, 2)\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \text{(i)}$$

$$\left.\begin{aligned}(0, 0, 1) &\text{ into } (1, 1, 0)\\ (1, 1, 1) &\text{ into } (2, 2, 0)\\ (-1, 1, 0) &\text{ into } (1, 3, -2)\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \text{(ii)}$$

$$\left.\begin{aligned}(1, 0) &\text{ into } (0, 1)\\ (0, 1) &\text{ into } (-1, 0)\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \text{(iii)}$$

Discuss the rank of these functions and calculate the inverse functional matrices if possible.

6.3. In Exercise 5, Section 6.6, it was stated that an orthonormal matrix $\mathbf{A}$ with det ($\mathbf{A}$) = −1 always has an eigenvalue of −1. Prove this contention and discuss whether it is necessary that the dimension of $\mathbf{A}$ be odd.

6.4. Let $\mathbf{A}$ and $\mathbf{B}$ be two like, square matrices. Prove that $\mathbf{AB}$ and $\mathbf{BA}$ have the same eigenvalues.

6.5. Apply the Cayley–Hamilton theorem to the skew matrix

$$\mathbf{A} = \begin{Bmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{Bmatrix}$$

6.6. Give a general discussion of the contents of Exercise 3, Section 3.7. *Hint*. Introduce the ranks of **A** and **B** and decompose the post-factor into column vectors.

6.7. Prove that the set of all n-by-n matrices which commute forms a group. How is this group related to the groups mentioned in Exercise 3, Section 6.5?

6.8. Prove that, if the square matrix **A** has eigenvalues λ_i ($\neq 0$), then $\mathbf{A}^{-1}$ will have the eigenvalues $1/\lambda_i$.

6.9. Verify that matrix **A** of Problem 5.5, considered as a functional matrix, rotates the whole of 3-space rigidly about the z-axis in such a way that the positive direction of the y-axis is carried into the positive direction of the x-axis. Show also that matrix **B** rotates 3-space about the y-axis in such a way that the positive direction of the x-axis falls along the positive direction of the z-axis.

Perform these rotations with (say) a matchbox so as to demonstrate that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \quad \text{(i)}$$

$$\mathbf{AB} \neq \mathbf{BA} \quad \text{(ii)}$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \quad \text{(iii)}$$

6.10. Express the linear function

$$\begin{aligned} x' &= x + y \\ y' &= \phantom{x + {}} y \end{aligned}$$

in matrix form as $X = \mathbf{A}Y$ and determine its eigenspace.

6.11. From classical vector analysis it is known that the area of the parallelogram determined by two vectors is equal to the product of their magnitudes and the sine of the angle between them. Prove that, if the area thus determined by the vectors V and W is a, then

$$a^2 = V_tVW_tW - V_tWV_tW = (\det(VW))^2$$

Hint. Use the fact that

$$V_tW = VW \,.\, \cos(V, W)$$

and

$$\sin^2(V, W) = 1 - \cos^2(V, W)$$

6.12. Let a 2-by-2 functional matrix $\mathbf{A} = (A_1A_2)$ be given. The two 2-dimensional unit vectors U_1 and U_2 define a unit square and transform as follows: U_1 into A_1 and U_2 into A_2: $(A_1A_2)(U_1U_2) = A_1A_2$).

Use this result to prove that the unit square is mapped into a parallelogram with an area equal to det (**A**). Show that (because of the linearity of the function) any area will be mapped into an area det (**A**) times larger.

6.13. Employ the result of Problem 6.12 to show that the function of Problem 6.10 leaves areas invariant.

6.14. Discuss and compare the linear vector functions

$$\mathbf{A}_1 = \begin{Bmatrix} 1 & -2 \\ 1 & 0 \end{Bmatrix} \quad \text{and} \quad \mathbf{A}_2 = \tfrac{1}{2}\mathbf{A}_1$$

and give a geometric interpretation of the equation $\det(k\mathbf{A}) = k^n \det(\mathbf{A})$ ($n = 2$) (*see* eqn. (3.6), p. 24).

6.15. Generalize the result of Problem 6.12 to three dimensions.

6.16. Show by direct matrix multiplication that matrices of the form

$$\begin{Bmatrix} \cos u & -\sin u \\ \sin u & \cos u \end{Bmatrix}$$

form a group under matrix multiplication. Discuss under what conditions a 2-by-2 matrix such as the one given above can generate a finite (cyclic and Abelian) group.

6.17. Prove that the set of matrices of the type

$$\begin{Bmatrix} \mathbf{Q} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{Bmatrix}$$

where **Q** is any r-by-r orthonormal matrix, and **I** is the $(n - r)$-by-$(n - r)$ unit matrix, form a group with respect to matrix multiplication, and show that such a group is a subgroup of the group of n-by-n orthonormal matrices.

6.18. Resolve the square matrix

$$\mathbf{A} = \begin{Bmatrix} 2 & 0 & 1 & 5 \\ 1 & -2 & 0 & 3 \\ 0 & 1 & 2 & 2 \\ 1 & -1 & 6 & 1 \end{Bmatrix}$$

into its symmetric and anti-symmetric components.

CHAPTER 7

Bilinear and Quadratic Forms

7.1. Definition of a Form

ANY combination of matrices and vector variables yielding a scalar is termed a *form*. In this book we shall be particularly concerned with bilinear and quadratic forms that occur so often in various disguises in geometry and applied mathematics in connexion with scalar quantities such as distance, angle, dissipated and stored energy, etc.

An expression

$$T = V_t \mathbf{A} W \quad . \quad . \quad . \quad . \quad (7.1)$$

where $\mathbf{A}$ is square and V and W are equi-dimensional variable vectors, is called a *bilinear form*. The adjective "bilinear" refers to the fact that eqn. (7.1) satisfies the linearity requirements (*see* Section 6.2, p. 61) for both vectors variables. For example,

$$(k_1 V_1 + k_2 V_2)_t \mathbf{A} W = k_1 V_{1t} \mathbf{A} W + k_2 V_{2t} \mathbf{A} W$$

and similarly for the variable W.

If T is invariant to an interchange of V and W, the form is said to be symmetric. This implies that the *form matrix* $\mathbf{A}$ is symmetric, since from the identity

$$V_t \mathbf{A} W = W_t \mathbf{A} V$$

we can conclude that $\mathbf{A}_t = \mathbf{A}$.

In the special case where V and W are one and the same vector, the form is called *quadratic* because

$$T = V_t \mathbf{A} V$$

is a second-degree polynomial in the coordinates of V.

Only the symmetric component of $\mathbf{A}$ need be considered (*see* Section 6.3, p. 63), since the quadratic form associated with a skew matrix vanishes identically, for when $\mathbf{A}$ is skew $\mathbf{A}_t = -\mathbf{A}$, and

$$\begin{aligned} V_t \mathbf{A} V &= (V_t \mathbf{A} V)_t \qquad \text{(because the form is a scalar)} \\ &= V_t \mathbf{A}_t V_{tt} \\ &= -V_t \mathbf{A} V \end{aligned}$$

which proves that $V_t \mathbf{A} V = 0$.

7.2. Transformation of the Form Matrix

When the coordinates of V and W are subject to a transformation $\mathbf{M}$ ($V' = \mathbf{M}V$, $W' = \mathbf{M}W$), the form matrix $\mathbf{A}$ will also transform—

$$\begin{aligned} T &= V_t \mathbf{A} W \\ &= (\mathbf{M}^{-1}V')_t \mathbf{A} (\mathbf{M}^{-1}W) \\ &= V'_t (\mathbf{M}_t^{-1} \mathbf{A} \mathbf{M}^{-1}) W' \\ &= V'_t \mathbf{A}' W' \end{aligned}$$

and we can formulate the next theorem.

Theorem 34. When the transformation $V' = \mathbf{M}V (W' = \mathbf{M}W)$ is applied to the coordinates of a bilinear or quadratic form, the form matrix $\mathbf{A}$ transforms according to the expression

$$\mathbf{A}' = \mathbf{M}_t^{-1} \mathbf{A} \mathbf{M}^{-1}$$

Comparing this with eqns. (6.3) and (6.5), and bearing in mind that the coordinate axes are covariant by convention, the form matrix is seen to be doubly covariant.

By referring back to Exercise 1, Section 6.1, we note that the metric of a space is in fact a form matrix.

Also, if we consider the orthonormal group of transformations, the transformation formula becomes

$$\mathbf{A}' = \mathbf{M} \mathbf{A} \mathbf{M}^{-1}$$

which is identical with that of Theorem 26, p. 65.

It has been pointed out that the form matrix $\mathbf{A}$ of a quadratic form is symmetric. Hence, by virtue of Theorem 30, p. 74, $\mathbf{A}$ has n real eigenvalues. Furthermore, since the transformations given in Theorems 26 and 34 are identical for orthonormal transformations, we can make use of Theorem 32, p. 75, to formulate Theorem 35.

Theorem 35. The form matrix of a quadratic form can always be reduced to a diagonal matrix by means of an orthonormal coordinate transformation.

After diagonalization the form matrix will be $\mathbf{A}_d$ and the quadratic form will be

$$T = V_t \mathbf{A}_d V = \lambda_1 v_1^2 + \lambda_2 v_2^2 + \ldots + \lambda_n v_n^2$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{A}$ (= the diagonal elements of $\mathbf{A}_d$).

The number (p) of positive eigenvalues is termed the *index* of T. If we denote the rank of $\mathbf{A}$ by r (= $r(\mathbf{A})$), then $n - r$ of the eigenvalues are zero, and $r - p$ are negative.

Quadratic forms are classified as indicated in Table 7.1.

Table 7.1
CLASSIFICATION OF QUADRATIC FORMS

p	r	Class
n	n	positive definite
r	$r < n$	positive
0	n	negative definite
0	$r < n$	negative
$< r$	$r \leqslant n$	indefinite

We note that positive and negative definite forms are equal to zero when, and only when, $V = 0$. For any other value of the vector variable the positive definite forms will always be positive and the negative definite forms negative.

Exercise. Transposing both members of the equation of Theorem 34, we get

$$(\mathbf{A}')_t = (\mathbf{M}_t^{-1}\mathbf{A}\mathbf{M}^{-1})_t = \mathbf{M}_t^{-1}\mathbf{A}_t\mathbf{M}_{tt}^{-1} = \mathbf{M}_t^{-1}\mathbf{A}\mathbf{M}^{-1} = \mathbf{A}'$$

Hence the symmetry of the form matrix is an invariant property under any non-singular coordinate transformation. Compare this result with that of the Exercise in Section 6.4.

7.3. Simultaneous Diagonalization of Two Form Matrices

Having diagonalized the form matrix **A** by means of an orthonormal transformation **N**, we can go a step further by applying a diagonal transformation **M** with the element $\sqrt{|\lambda_i|}$ in position *ii* on the main diagonal. The corresponding element of $\mathbf{M}^{-1}$ (which when **M** is a diagonal matrix is equal to $\mathbf{M}_t^{-1}$) is $1\sqrt{|\lambda_i|}$. When an eigenvalue is zero any non-zero number can be used in the corresponding place in **M**; it will subsequently cancel out when the transformation is applied to $\mathbf{A}_d$.

Transforming $\mathbf{A}_d$ by **M** results in a *normalized* diagonal matrix—

$$\begin{aligned}\mathbf{A}_{dn} &= \mathbf{M}_t^{-1}\mathbf{A}_d\mathbf{M}^{-1}\\ &= \mathbf{M}_t^{-1}(\mathbf{N}_t^{-1}\mathbf{A}\mathbf{N}^{-1})\mathbf{M}^{-1}\\ &= (\mathbf{MN})_t^{-1}\mathbf{A}(\mathbf{MN})^{-1}\end{aligned}$$

with elements +1, −1 and 0 along the main diagonal according to whether the corresponding eigenvalues are positive, negative or zero. It should be carefully noted that the normalizing process hinges on the double covariance of the form matrix.

If **A** is positive (negative) definite, its normalized diagonal form will be equal to **I** (**−I**), and it will therefore be invariant to any

subsequent orthonormal transformation **K**, as is immediately apparent from the expression

$$\mathbf{K}(\pm\mathbf{I})\mathbf{K}_t = \pm\mathbf{K}\mathbf{K}_t = \pm\mathbf{I}$$

Exercise 1. In the general case, the normalized diagonal form matrix can be written

$$\mathbf{A}_{dn} = \begin{Bmatrix}\mathbf{I} & \mathbf{O} & \mathbf{O}\\ \mathbf{O} & -\mathbf{I} & \mathbf{O}\\ \mathbf{O} & \mathbf{O} & \mathbf{O}\end{Bmatrix}$$

This form is invariant (as can be readily verified) to orthonormal transformations of the type

$$\mathbf{K} = \begin{Bmatrix}\mathbf{P} & \mathbf{O} & \mathbf{O}\\ \mathbf{O} & \mathbf{Q} & \mathbf{O}\\ \mathbf{O} & \mathbf{O} & \mathbf{R}\end{Bmatrix}$$

where the square submatrices **P**, **Q** and **R** are orthonormal, and the partitioning of **K** is such that the product $\mathbf{K}\mathbf{A}_{dn}\mathbf{K}_t$ is defined.

Suppose now that two quadratic forms T_1 and T_2 with form matrices **A** and **B**, respectively, are given. Further, let us assume that **A** is positive (negative) definite. We have shown that it is always possible to diagonalize **A** by means of an orthonormal transformation **N**, and then to normalize $\mathbf{A}_d$ by applying a diagonal transformation **M**. The resultant matrix is

$$\mathbf{A}_{dn} = (\mathbf{MN})_t^{-1}\mathbf{A}(\mathbf{MN})^{-1} = \mathbf{I}\,(-\mathbf{I})$$

which, as we have seen, is invariant to any subsequent application of an orthonormal transformation.

Applying the transformation **MN** to the form **B** gives us

$$\mathbf{B}' = (\mathbf{MN})_t^{-1}\mathbf{B}(\mathbf{MN})^{-1}$$

B′ being symmetric can be reduced to diagonal form by means of an orthonormal transformation **K** (which leaves $\mathbf{A}_{dn}$ invariant).

Hence, we have proved the next theorem.

Theorem 36. Two form matrices **A** and **B**, where **A** is positive (negative) definite, can be simultaneously reduced to normalized diagonal and diagonal form, respectively, by means of a transformation **KMN**, where **K** and **N** are orthonormal and **M** is a diagonal matrix.

From the preceding discussion the procedure for simultaneously diagonalizing two matrices **A** and **B** is clear. We first determine **N** as described in Section 6.6, p. 67, and then compute the normalizing transformation **M**. After transforming **B** into **B′** by means of the

matrix **MN**, we calculate the eigenvalues of **B**′ and the associated diagonalizing transformation **K**.

There is, however, a short-cut procedure for computing the final quadratic form without having to determine **N**, **M** or **K**.

Let T_1 and T_2 be the quadratic forms associated with **A** and **B**, respectively.

The quadratic form

$$T_2 - kT_1 = V_t(\mathbf{B} - k\mathbf{A})V$$

transforms into

$$\begin{aligned} T_2 - kT_1 &= W_t(\mathbf{B}_d - k\mathbf{A}_{dn})W \\ &= (\lambda_1 - k)w_1^2 + (\lambda_2 - k)w_2^2 + \ldots + (\lambda_n - k)w_n^2 \end{aligned}$$

in which equation

$$\mathbf{B}_d - k\mathbf{A}_{dn} = (\mathbf{KMN})_t^{-1}(\mathbf{B} - k\mathbf{A})(\mathbf{KMN})^{-1}$$

Hence, $\det(\mathbf{B}_d - k\mathbf{A}_{dn}) = (\lambda_1 - k)(\lambda_2 - k)\ldots(\lambda_n - k)$

$$= \det{}^2(\mathbf{KMN})^{-1} . \det(\mathbf{B} - k\mathbf{A})$$

where $\det^2(\mathbf{KMN})^{-1}$ is independent of k.

This means that the determinants

$$\det(\mathbf{B}_d - k\mathbf{A}_{dn}) = (\lambda_1 - k)(\lambda_2 - k)\ldots(\lambda_n - k)$$

and $\det(\mathbf{B} - k\mathbf{A})$ are proportional nth-degree polynomials in k and have, therefore, the same roots $k = \lambda_1, \ldots, \lambda_n$.

To calculate the constants $\lambda_1, \ldots, \lambda_n$, we seek the roots of the equation $\det(\mathbf{B} - k\mathbf{A}) = 0$.

When **A** is positive (negative) definite $\mathbf{A}^{-1}$ exists and the above equation has roots which are the eigenvalues of the matrix $\mathbf{A}^{-1}\mathbf{B}$.

Exercise 2. To illustrate the method of simultaneous diagonalization, let

$$\mathbf{A} = \begin{Bmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{Bmatrix}$$

which we know from Exercise 2, Section 6.6, to be positive definite, and let

$$\mathbf{B} = \begin{Bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{Bmatrix}$$

We find that $\det(\mathbf{B} - k\mathbf{A}) = -162k^3 + 43k^2 + 17k = 0$

whose roots are $k = 0{\cdot}48$, $-0{\cdot}22$ and 0.

T_1 thus reduces to $w_1^2 + w_2^2 + w_3^2$, and T_2 to $0{\cdot}48w_1^2 - 0{\cdot}22w_2^2$.

7.4. Equivalence Relations and Equivalence Classes

A correspondence or equivalence between the elements $a, b, c, \ldots$ of a set is said to be an *equivalence relation* (an R.S.T. relation) if the following conditions are satisfied. The correspondence must be—

(i) *Reflexive*: any element is equivalent to itself.

(ii) *Symmetric*: if a is equivalent to b, then b is equivalent to a.

(iii) *Transitive*: if a is equivalent to b, and b is equivalent to c, it follows that a is equivalent to c.

The establishment of an equivalence relation separates the members of a set into mutually exclusive subsets called *equivalence classes*.

Let $a_1, b_1, c_1, \ldots$ belong to class C_1, and $a_2, b_2, c_2, \ldots$, to class C_2. It is easy to prove that these two classes are either identical or have no element in common. If one element a_1 (say) of class C_1 is equivalent to an element a_2 (say) of class C_2, the transitive property of the relation will cause a_1 to be equivalent to all the elements of C_2. Furthermore, the symmetry of the equivalence makes a_2 equivalent to a_1, and hence transitively to all the elements of class C_1.

Thus, two equivalence classes cannot have an element in common without becoming identical, and are therefore mutually exclusive.

To give an instance of a system of such classes, consider the set of all (symmetric) n-by-n form matrices **A**, **B**, **C**, ...

We shall define **A** to be equivalent to **B** if there exists a transformation belonging to the group G_{on} of n-dimensional orthonormal transformations, such that

$$\mathbf{B} = \mathbf{M}_t^{-1}\mathbf{A}\mathbf{M}^{-1} = \mathbf{M}\mathbf{A}\mathbf{M}^{-1} \quad . \quad . \quad . \quad (7.2)$$

This correspondence between **A** and **B** (which is sometimes called *congruence*) is an R.S.T.-relation.

It is reflexive, since **A** is equivalent to itself,

$$\mathbf{A} = \mathbf{I}\mathbf{A}\mathbf{I}^{-1}$$

where **I** is the (orthonormal) unit element of the group G_{on}.

It is also symmetric. Pre- and post-multiplication of eqn. (7.2) by $\mathbf{M}^{-1}$ and **M**, respectively, yield

$$\mathbf{A} = \mathbf{M}^{-1}\mathbf{B}\mathbf{M} = (\mathbf{M}^{-1})\mathbf{B}(\mathbf{M}^{-1})^{-1}$$

i.e. **B** is equivalent to **A**, where $\mathbf{M}^{-1}$, according to the fourth group axiom, also belongs to G_{on}.

Finally, the correspondence is transitive because the equivalence of **A** and **B**,

$$\mathbf{B} = \mathbf{M}\mathbf{A}\mathbf{M}^{-1} \quad . \quad . \quad . \quad . \quad (7.3)$$

and the equivalence of **B** and **C**,

$$\mathbf{C} = \mathbf{NBN}^{-1} \quad . \quad . \quad . \quad . \quad . \quad (7.4)$$

implies (by substituting eqn. (7.3) in eqn. (7.4)) that **A** is equivalent to **C**:

$$\mathbf{C} = \mathbf{N}(\mathbf{MAM}^{-1})\mathbf{N}^{-1} = (\mathbf{NM})\mathbf{A}(\mathbf{NM})^{-1}$$

where **NM** is a member of G_{on}, since both **M** and **N** belong to G_{on} and the group is closed.

We shall see in Section 8.1 that the above classification of quadratic forms admits of a very simple geometric interpretation in 2- and 3-space.

PROBLEMS

7.1. Diagonalize the quadratic forms

$$x^2 - 2xy + y^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (i)$$

$$z^2 - 2xy \quad . \quad . \quad . \quad . \quad . \quad . \quad (ii)$$

7.2. Write the quadratic form $T = x^2 - 6xy - y^2$ in the form $T = \mathbf{X}_t\mathbf{A}\mathbf{X}$, and diagonalize it.

7.3. Apply the linear transformation

$$\begin{aligned} x &= x' + y' + z' \\ y &= \phantom{x' + {}} y' + z' \\ z &= x' - y' \end{aligned}$$

to the quadratic form $z^2 - 2xy$ by means of the equation of transformation given in Theorem 34 and test the result by direct substitution.

7.4. Reduce the expressions

$$z^2 - 4z - 2xy + 6x - 2y \quad . \quad . \quad . \quad . \quad . \quad (i)$$

$$x^2 - 6x + y^2 \quad + 6y - 2xy \quad . \quad . \quad . \quad . \quad . \quad (ii)$$

by means of coordinate translations $x = x' + a$, $y = y' + b$, and $z = z' + c$, so as to eliminate the linear terms.

7.5. Discuss the conditions for positive definiteness for the quadratic form $ax^2 + 2bxy + cy^2$, and show that the two eigenvalues of the form matrix are always real.

7.6. Test the assertion in the first paragraph of Section 7.3 that the application of the transformation **M** reduces $\mathbf{A}_s$ to normalized diagonal form.

7.7. Prove that the normalized diagonal matrix of a positive form is invariant to any orthonormal transformation of the type

$$\begin{Bmatrix} \mathbf{Q} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{Bmatrix}$$

where **Q** is an r-by-r orthonormal matrix, and **I** is the $(n - r)$-by-$(n - r)$ unit matrix.

7.8. Let a, b and c be the coefficients of the quadratic form given in Problem 7.5. Prove by direct calculation that $a + c$ and $ac - b^2$ are invariants under the group of 2-by-2 orthonormal transformations.

7.9. The kinetic energy of a particle, with mass m and moving with velocity V, is $T = \frac{1}{2}mV_tV$. Show that T is invariant to any rigid rotation of an orthogonal coordinate system.

7.10. Prove the statement in the last paragraph of Section 7.3 that the roots k_i of the equation $\det(\mathbf{B} - k\mathbf{A}) = 0$ are the eigenvalues of $\mathbf{A}^{-1}\mathbf{B}$.

7.11. Show that the relationship between the sets of vectors A and B introduced in Section 2.4, p. 10, is an R.S.T.-relation. What do the corresponding equivalence classes represent geometrically?

CHAPTER 8

Application of Matrices to Geometry and Mechanics

8.1. Central Quadric Surfaces

WE are now equipped to carry out a complete analysis of algebraic surfaces (and curves) of the second degree. To save space, however, we shall limit ourselves to a brief discussion of central quadric surfaces, i.e. surfaces that can be expressed by an equation of the type

$$ax^2 + by^2 + cz^2 + 2dyz + 2exz + 2fxy = k \qquad . \quad (8.1)$$

where k is a positive constant.

The adjective "central" refers to the fact that the origin $O = (0, 0, 0)$ is the centre of any surface of this type. If $P = (x_0, y_0, z_0)$ is a point on the surface, then $Q = (-x_0, -y_0, -z_0)$, which is symmetric to P with respect to O, will also satisfy eqn. (8.1). PQ is termed a diameter of the quadric.

In this Section we shall consider only Cartesian coordinates and orthonormal coordinate transformations (i.e. rigid rotations of the coordinate system).

Eqn. (8.1) is readily generalized to n-space, and can be written very concisely as a quadratic form

$$X_t\mathbf{A}X = k \qquad . \quad . \quad . \quad . \quad (8.2)$$

where $\mathbf{A}$ is symmetric and therefore has n real eigenvalues. By virtue of Theorem 32, p. 75, it is always possible to diagonalize the form matrix, which, in terms of eqn. (8.1), means that all product terms vanish, leaving only the squares.

The various possibilities in Euclidean 3-space are set out in Table 8.1.

When one of the eigenvalues is zero, the corresponding surface (if it exists) degenerates into an elliptic or hyperbolic cylinder. If two eigenvalues vanish, the equation will represent two parallel planes perpendicular to one of the coordinate axes.

In the special case $k = 0$, we note that, if a point $P = (x_0, y_0, z_0)$ satisfies the equation

$$X_t\mathbf{A}X = 0$$

then any point $Q = (tx_0, ty_0, tz_0)$, where t is any real number, will also be a solution. Thus the corresponding surface is a cone. The indefinite forms will represent proper elliptic cones, whereas positive-definite or negative-definite forms will only represent the origin (0, 0, 0).

Table 8.1

PROPER QUADRIC SURFACES IN 3-SPACE

λ_1	λ_2	λ_3	Class	Surface
pos.	pos.	pos.	positive definite	Ellipsoid. When two λ's are equal, a rotational ellipsoid. When all three λ's are equal, a sphere.
pos.	pos.	neg.	indefinite	Hyperboloid with one sheet. When the two pos. λ's are equal, a rotational hyperboloid ("cooling tower").
pos.	neg.	neg.	indefinite	Hyperboloid with two sheets. When the two neg. λ's are equal, a rotational hyperboloid.
neg.	neg.	neg.	negative definite	The equation represents nothing.

To return to the case where $k > 0$, we see that two quadrics with identical matrices, but different values of k, are *homothetic* (i.e. the one surface can be derived from the other by multiplication from the origin). If (x_0, y_0, z_0) is a point on the surface corresponding to the constant k_1, then $\{\sqrt{(k_2/k_1)}\}(x_0, y_0, z_0)$ is a point of the surface associated with k_2. This result stems from the bilinearity of quadratic forms.

Interpreted geometrically, the classification of form matrices given in Section 7.4, p. 87, can readily be visualized. For a constant value of k, equivalent matrices correspond to the same quadric surface referred to different Cartesian coordinate systems. It is obvious, therefore, that the classes must be discrete.

In 2-space a non-degenerate quadric is either an ellipse (with the circle as a special case) or a hyperbola. Degenerate curves are two parallel lines or, when $k = 0$, two intersecting lines through the origin.

Exercise 1. In order to analyse the quadric

$$5x^2 + 2y^2 + 2z^2 - 4xy - 4xz - 8yz = 108$$

which can be written in matrix form as

$$\{x \;\; y \;\; z\}\begin{bmatrix} 5 & -2 & -2 \\ -2 & 2 & -4 \\ -2 & -4 & 2 \end{bmatrix}\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = 108$$

we must first find the roots of the reduction determinant

$$\det(\mathbf{R}(\lambda)) = -\lambda^3 + 9\lambda^2 - 108 = -(\lambda + 3)(\lambda - 6)^2 = 0$$

The eigenvector corresponding to the eigenvalue -3 is found to be

$$V_1 = \begin{Bmatrix} 1 \\ 2 \\ 2 \end{Bmatrix}$$

From Theorem 28, p. 69, we know that the eigenvectors of the double eigenvalue $+6$ span a 2-space and are orthogonal to V_1. They must lie in the plane

$$x + 2y + 2z = 0$$

As one of them we choose (arbitrarily)

$$V_2 = \begin{Bmatrix} 0 \\ 1 \\ -1 \end{Bmatrix}$$

and as the other we calculate the vector which is perpendicular to both V_1 and V_2; this gives us

$$V_3 = \begin{Bmatrix} -4 \\ 1 \\ 1 \end{Bmatrix}$$

From these three orthogonal vectors we compute a diagonalizing transformation matrix by normalizing the columns of the matrix

$$(V_1 V_2 V_3)$$

This leads to the matrix

$$\mathbf{M} = \begin{Bmatrix} 1/3 & 0 & -4\sqrt{2}/6 \\ 2/3 & \sqrt{2}/2 & \sqrt{2}/6 \\ 2/3 & -\sqrt{2}/2 & \sqrt{2}/6 \end{Bmatrix}$$

which transforms $\mathbf{A}$ into

$$\mathbf{A}' = \begin{Bmatrix} -3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{Bmatrix}$$

The corresponding quadratic form is

$$-3x^2 + 6y^2 + 6z^2 = 108$$

or

$$-x^2/6^2 + y^2/(3\sqrt{2})^2 + z^2/(3\sqrt{2})^2 = 1$$

which represents a rotational hyperboloid with one sheet, the x-axis being the axis of rotation.

Exercise 2. The form

$$X_t \mathbf{I} X = x^2 + y^2 + z^2 = k \qquad (> 0)$$

represents a sphere with its centre in the origin and radius $R = \sqrt{k}$. $\lambda = 1$ is a triple root of $\det(\mathbf{R}(\lambda)) = 0$, and any vector is an eigenvector of the form matrix $\mathbf{I}$. Thus, any orthonormal transformation will leave the form invariant—

$$\mathbf{I}' = \mathbf{M}\mathbf{I}\mathbf{M}^{-1} = \mathbf{M}\mathbf{M}^{-1} = \mathbf{I}$$

The geometric significance of this last result is quite clear and should be compared with the proof in Section 7.3, p. 85, that a non-indefinite diagonalized and normalized matrix $\mathbf{A}_{dn}$ is invariant to orthonormal transformations.

Exercise 3. An equation such as $2xy = 9$ is a quadratic form with the form matrix

$$\mathbf{A} = \begin{Bmatrix} 0 & 1 \\ 1 & 0 \end{Bmatrix}$$

which was diagonalized by means of an orthonormal transformation in Exercise 3, Section 6.8.

The reduced (or canonical) form of the equation is

$$x^2 - y^2 = 9$$

or

$$(x/3)^2 - (y/3)^2 = 1$$

representing an equilateral hyperbola.

8.2. Relative Motion

Consider two right-handed orthonormal coordinate systems S and S'. S' is assumed to move relative to S, and we now set ourselves the task of finding the relationship between the velocities and accelerations of a point expressed in either S- or S'-coordinates.

In S, the position vector of point P is R, and in S', R'. Expressed in matrix notation, the connexion between the coordinates of R and R' is

$$R = Q + \mathbf{M}R' \qquad (8.3)$$

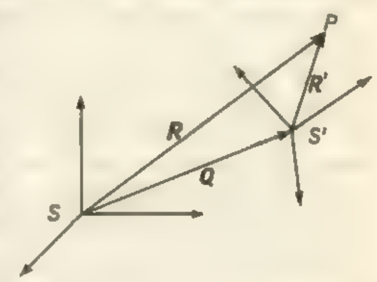


FIG. 8.1. ABSOLUTE AND RELATIVE MOTION

where $\mathbf{M}$ is an orthonormal matrix which depends upon the position of S' relative to S, and the components of which are therefore continuous functions of the scalar time parameter t.

Differentiation with respect to time gives us

$$\dot{R} = \dot{Q} + \dot{\mathbf{M}}R' + \mathbf{M}\dot{R}' \qquad (8.4)$$

(we employ the Newtonian dot notation).

$\dot{R}$ is the velocity of P relative to S; $\dot{R}'$, the velocity of P relative to S'; and $\dot{Q} + \dot{\mathbf{M}}R'$, the velocity relative to S of the fixed point in S' which is instantaneously coincident with P. It is standard practice to call these three terms *absolute*, *relative* and *transport* velocities, respectively.

We shall now study the term $\dot{\mathbf{M}}R'$ more carefully.

As S' moves rigidly relative to S, the components of $\mathbf{M}$ will vary, the only condition imposed on them being that $\mathbf{M}$ shall at all times be orthonormal. The relation

$$\mathbf{M}\mathbf{M}_t = \mathbf{M}_t\mathbf{M} = \mathbf{I} \qquad (8.5)$$

is thus independent of time, and by differentiation we find

$$\dot{\mathbf{M}}\mathbf{M}_t + \mathbf{M}\dot{\mathbf{M}}_t = 0 \quad . \quad . \quad . \quad (8.6)$$

or

$$\dot{\mathbf{M}} = -\mathbf{M}\dot{\mathbf{M}}_t\mathbf{M} \quad . \quad . \quad . \quad . \quad (8.7)$$

Applying Theorem 15, p. 26, to eqn. (8.7), we get

$$\begin{aligned} \det(\dot{\mathbf{M}}) &= -\det(\mathbf{M})\det(\dot{\mathbf{M}}_t)\det(\mathbf{M}) \\ &= -\det(\dot{\mathbf{M}}_t) \qquad \text{(because } \det(\mathbf{M}) = 1\text{)} \\ &= -\det(\dot{\mathbf{M}}) \qquad \text{(because } \det(\mathbf{A}_t) = \det(\mathbf{A})\text{)} \end{aligned}$$

from which we infer (by comparing the first and last members of the equation) that

$$\det(\dot{\mathbf{M}}) = 0$$

for all values of t.

$\dot{\mathbf{M}}$ will thus always have the eigenvalue zero, and the corresponding eigenvector $\boldsymbol{R}_0'$ will be instantaneously at rest relative to S'.

Since $\dot{\mathbf{M}}\boldsymbol{R}'$ is a linear vector function and $\dot{\mathbf{M}}\boldsymbol{R}_0' = \boldsymbol{0}$, $\dot{\mathbf{M}}\boldsymbol{R}'$ will depend solely on the component of $\boldsymbol{R}'$ perpendicular to $\boldsymbol{R}_0'$. Also, the magnitude of $\dot{\mathbf{M}}\boldsymbol{R}'$ will be proportional to the distance from the terminal point of $\boldsymbol{R}'$ to the eigendirection determined by $\boldsymbol{R}_0'$. Furthermore, $\mathbf{M}\boldsymbol{R}'$ and $\dot{\mathbf{M}}\boldsymbol{R}'$ (the velocity relative to S of the vector $\boldsymbol{R}'$ which is fixed in S') are orthogonal, since

$$\begin{aligned} (\dot{\mathbf{M}}\boldsymbol{R}')_t(\mathbf{M}\boldsymbol{R}') &= \boldsymbol{R}_t'\dot{\mathbf{M}}_t\mathbf{M}\boldsymbol{R}' && \text{(by Theorem 14, p. 26)} \\ &= -\boldsymbol{R}_t'\mathbf{M}_t\dot{\mathbf{M}}\boldsymbol{R}' && \text{(by eqn. (8.6))} \\ &= -\boldsymbol{R}_t'\dot{\mathbf{M}}_t\mathbf{M}\boldsymbol{R}' && \text{(by transposition)} \\ &= 0 && \text{(by comparing the second and fourth terms of this string of identities)} \end{aligned}$$

Also, $\dot{\mathbf{M}}\boldsymbol{R}'$ is orthogonal to $\mathbf{M}\boldsymbol{R}_0'$ (the eigenvector of $\dot{\mathbf{M}}$ referred to S)—

$$\begin{aligned} (\dot{\mathbf{M}}\boldsymbol{R}')_t(\mathbf{M}\boldsymbol{R}_0') &= \boldsymbol{R}_t'\dot{\mathbf{M}}_t\mathbf{M}\boldsymbol{R}_0' && \text{(by Theorem 14)} \\ &= -\boldsymbol{R}_t'\mathbf{M}_t\dot{\mathbf{M}}\boldsymbol{R}_0' && \text{(by eqn. (8.6))} \\ &= -\boldsymbol{R}_t'\mathbf{M}_t\boldsymbol{0} && \text{(since } \boldsymbol{R}_0' \text{ is an eigenvector of } \dot{\mathbf{M}} \text{ corresponding to } \lambda = 0\text{)} \\ &= 0 \end{aligned}$$

The reader familiar with classical kinematics will recognize that the instantaneous motion of S' relative to S is a rotation about the axis determined by $\boldsymbol{R}_0'$. A point $\boldsymbol{R}'$ in S' moves relative to S with a velocity

$$\dot{\boldsymbol{Q}} + \boldsymbol{\omega} \times \boldsymbol{R}'$$

(reverting for a moment to classical vector notation), where $\boldsymbol{\omega} \times \boldsymbol{R}'$ is orthogonal to both $\boldsymbol{\omega}$ (which lies along $\boldsymbol{R}_0'$) and $\boldsymbol{R}'$, and proportional to the distance from the terminal of $\boldsymbol{R}'$ to the axis of rotation.

To find the acceleration of P, we differentiate a second time and obtain

$$\ddot{\boldsymbol{R}} = (\ddot{\boldsymbol{Q}} + \ddot{\mathbf{M}}\boldsymbol{R}') + \mathbf{M}\ddot{\boldsymbol{R}}' + 2\dot{\mathbf{M}}\dot{\boldsymbol{R}}' \quad . \quad . \quad (8.8)$$

$\ddot{\boldsymbol{R}}$ is the absolute acceleration of P relative to S; $\ddot{\boldsymbol{Q}} + \ddot{\mathbf{M}}\boldsymbol{R}'$ is the acceleration in S of the fixed point in S' instantaneously coincident with $\boldsymbol{R}'$ (transport acceleration); and $\ddot{\boldsymbol{R}}'$ is the acceleration of $\boldsymbol{R}'$ in S' (relative acceleration). $\ddot{\boldsymbol{R}}'$ is pre-multiplied by $\mathbf{M}$ to transform it into the S-system.

Finally, the term $2\dot{\mathbf{M}}\dot{\boldsymbol{R}}'$ is Coriolis' acceleration. This last term of eqn. (8.8) will vanish under three conditions—

(i) When S' is instantaneously at rest in S ($\dot{\mathbf{M}} = \mathbf{0}$).
(ii) When P is instantaneously at rest in S' ($\dot{\boldsymbol{R}}' = \boldsymbol{0}$).
(iii) When $\dot{\boldsymbol{R}}'$ lies along the instantaneous axis of rotation (the eigendirection of $\dot{\mathbf{M}}$).

Note. This section is a typical example of an engineer rushing in where mathematicians would fear to tread, at least before they had studied the ground more carefully. We have optimistically differentiated matrices according to the well-known rules for scalar functions without first satisfying ourselves that such an operation is in fact permissible. The operator d/dt is linear, however, and it is thus a reasonable assumption, which is readily verified, that we differentiate matrix products with impunity according to the classical rules, provided the order of the matrix factors is not altered.

8.3. Tensor of Inertia

As an example of the application of matrix algebra to classical mechanics, we shall now consider a rigid body, one point of which is fixed by means of a frictionless joint.

In using the word *tensor* we have anticipated the contents of Chapter 12. For the time being, however, the reader can think of a tensor as a matrix representing a class of orthonormally equivalent (congruent) form matrices such as were discussed in Section 7.4, p. 87.

The moment of inertia of a heavy body about an axis through a fixed point O is defined as the scalar sum of the mass particles of which the body consists, each multiplied by the square of its distance from the axis about which the moment is being calculated.

Choosing O as the origin of our system of reference, $\boldsymbol{R}$ as the

position vector of the mass particle m, and L as a unit vector defining the axis of rotation, the moment of inertia about OQ is

$$J \stackrel{D}{=} \Sigma m(PQ)^2 \qquad (8.9)$$

According to Pythagoras, $(PQ)^2 = (OP)^2 - (OQ)^2$, which in the notation of classical vector analysis can be written

$$(PQ)^2 = R^2 - (L \,.\, R)^2 \qquad (8.10)$$

where $L \,.\, R = OQ$, since L is a unit vector.

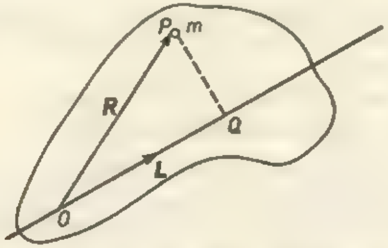


FIG. 8.2. MOMENT OF INERTIA OF RIGID BODY

Substituting eqn. (8.10) in eqn. (8.9), and expressing the scalar products in matrix notation, we get

$$J = \Sigma m R_t R - \Sigma m(L_t R R_t L) \qquad (8.11)$$

Since L is a unit vector, $L_t L = L_t \mathbf{I} L = 1$, and by inserting the scalar factor mR_tR between L_t and $\mathbf{I}$, eqn. (8.11) can be rearranged as follows—

$$\begin{aligned} J &= \Sigma L_t(mR_tR\mathbf{I})L - \Sigma L_t(mRR_t)L \\ &= L_t(\Sigma(mR_tR\mathbf{I} - mRR_t))L \\ &= L_t \begin{Bmatrix} \Sigma m(y^2+z^2) & -\Sigma mxy & -\Sigma mxz \\ -\Sigma mxy & \Sigma m(x^2+z^2) & -\Sigma myz \\ -\Sigma mxz & -\Sigma myz & \Sigma m(x^2+y^2) \end{Bmatrix} L \\ &= L_t \begin{Bmatrix} J_x & -D_{xy} & -D_{xz} \\ -D_{xy} & J_y & -D_{yz} \\ -D_{xz} & -D_{yz} & J_z \end{Bmatrix} L \\ &= L_t \mathbf{T} L \qquad (8.12) \end{aligned}$$

J_x, J_y and J_z are the moments of inertia about the coordinate axes; the D's are the products of inertia (or deviation moments) of the body with respect to the coordinate planes; and $\mathbf{T}$ is the *tensor of inertia* of the body corresponding to the point O.

$\mathbf{T}$ is symmetric and will therefore always have three real eigenvalues, J_1, J_2 and J_3, which are called the *principal moments of inertia.* The associated eigendirections are the principal axes of the body through point O, and when we swing the coordinate axes round to coincide with these directions (which are orthogonal), $\mathbf{T}$ transforms into

$$\mathbf{T}' = \begin{Bmatrix} J_1 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & J_3 \end{Bmatrix}$$

From physical considerations it is obvious that the principal moments of inertia can never be negative. Hence, the central quadric surface associated with the symmetric matrix $\mathbf{T}$ is an ellipsoid (the *ellipsoid of inertia*), the principal semi-axes of which are equal to the reciprocals of the square roots of the principal moments of inertia.

Exercise 1. As a beautiful example of the way in which matrix algebra, assisted by geometry, enables us to visualize and solve problems that would otherwise require laborious calculations, we shall discuss the moments of inertia of a homogeneous cube about lines through its centre.

For reasons of symmetry, the deviation moments are zero when we place the coordinate system with its axes perpendicular to the faces of the cube. For the same reason the moments of inertia about these axes are equal.

Hence, the ellipsoid of inertia is a sphere, and the moment of inertia of the cube about any line through its centre is a constant.

Exercise 2. The property of a form matrix that its trace (i.e. the sum of the elements along its main diagonal) is invariant to orthonormal transformations can now be interpreted mechanically: the sum of the moments of inertia of a rigid body about three perpendicular axes through a fixed point is a constant.

This can easily be proved directly by compiling the trace of the matrix of eqn. (8.12)—

$$J_x + J_y + J_z = 2\Sigma m(x^2 + y^2 + z^2) = 2\Sigma m R_t R$$

which is seen to be independent of L.

PROBLEMS

8.1. Determine the centre and the principle axes (the eigendirections) of the quadric surface

$$8x^2 + 11y^2 + 8z^2 + 4yz + 8zx - 4xy - 16x - 4y - 8z - 4 = 0$$

Method. Translate the coordinate system so as to eliminate the linear terms; then determine the eigenvalues and eigenvectors of the form matrix.

8.2. Discuss the central quadric surface $xy = z^2$.

8.3. Give a geometric interpretation of the normalizing process described in Section 7.3, p. 84.

8.4. Explain in geometric terms why it is always possible to diagonalize two quadratic forms simultaneously, provided that at least one of them is positive definite. Is this requirement necessary?

8.5. Show that eqns. (8.3), (8.4) and (8.8) remain invariant in functional form when we make S' the "absolute" and S the "relative" coordinate system. *Hint.* Note that pre-multiplication by $\mathbf{M}$ transforms vectors from S' into S, and (since $\mathbf{M}$ is orthonormal) $\mathbf{M}_t = \mathbf{M}^{-1}$ will transform vectors from S into S'.

8.6. Prove that the velocity relative to S' of the fixed point in S instantaneously coincident with P is equal to the negative of the velocity relative to S of the fixed point in S' instantaneously coincident with P. *Hint.* Make use of eqn. (8.6).

8.7. Verify the expanded form of the expression $\Sigma(m\boldsymbol{R}_t\boldsymbol{R}\mathbf{I} - m\boldsymbol{R}\boldsymbol{R}_t)$, given in eqn. (8.12).

8.8. Calculate the tensor of inertia at O of a composite body consisting of six heavy particles m placed in positions $(\pm 1, 0, 0)$, $(0, \pm 1, 0)$ and $(0, 0, \pm 1)$.

8.9. Use the tensor calculated in Problem 8.8 to determine the moment of inertia of the body about the axis determined by the vector $(1, 1, 1)$. Compare this result with the value obtained by direct calculation.

8.10. Repeat Problems 8.8 and 8.9 omitting the particles on the z-axis.

8.11. Compute the tensor of inertia at O of a composite body consisting of three heavy particles m placed at the three points $(1, 0, 0)$, $(0, 1, 0)$ and $(-1, -1, 1)$. Find the principal axes of the body.

8.12. Compute the tensor of inertia at O of a heavy particle m placed at point $(1, 0, 0)$. Sketch the (degenerate) ellipsoid of inertia of the body.

8.13. Choose a suitable Cartesian coordinate system and compute the tensor of inertia of a thin, heavy rod of mass M and length $2l$ about its centre. Determine the corresponding ellipsoid of inertia.

8.14. Calculate the tensor of inertia at the centre of a homogeneous circular disc of mass M and radius R.

8.15. Discuss the central quadric surface corresponding to a form matrix with (i) one eigenvalue equal to zero, and (ii) two eigenvalues equal to zero.

CHAPTER 9

Linear Programming

9.1. The Problem of Linear Programming

DURING recent years, linear algebra has found interesting and unexpected applications in industry and commerce under the name of *linear programming* (L.P.), which originated in military *operations research*. In this chapter we shall confine ourselves to a brief exposition of the problem of L.P., followed by a cursory discussion of the mathematical aspects of the subject.

Suppose that a certain industrial undertaking manufactures n different items, the quantities of which are x_i $(\geqslant 0)$ $(i = 1, 2, \ldots, n)$, and that the x_i can be sold at a net profit c_i on a market which does not exhibit any saturation effects. Let us assume further that the x_i are not completely independent of each other. For instance, x_1 may be produced from the waste products of item x_2; hence x_1 and x_2 are linked by an expression of the form

$$a_1x_1 + a_2x_2 = 0$$

Interdependence might also occur if x_1 and x_2 were manufactured by the same machine (a loom, for example, can be adjusted to produce rolls of cloth of different width).

The problem of L.P. is now to determine a production schedule which will take these limitations into consideration and at the same time yield a maximum profit. Linear programming has also been applied to transportation problems, in which case the aim of the analysis is to reduce costs to a minimum; the L.P. technique is, however, *mutatis mutandis*, precisely the same in each case.

Stripped of all its trappings, the mathematical formulation of the problem of linear programming is as follows.

Determine n non-negative numbers x_i such that the $m\,(< n)$ equations of constraint,

$$\left.\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + \ldots + & a_{1n}x_n & = & b_1 \\ \cdot & & \cdot & \cdots & \cdot & & \cdot \\ \cdot & & \cdot & \cdots & \cdot & & \cdot \\ a_{m1}x_1 & + & a_{m2}x_2 & + \ldots + & a_{mn}x_n & = & b_m \end{array}\right\} \quad . \quad . \quad (9.1)$$

are satisfied, and the scalar product,

$$f = \sum_{i}^{n} c_i x_i \overset{D}{=} c_i x_i \qquad . \quad . \quad . \ (9.2)$$

attains its maximum value.

Expressed in terms of compound matrices, eqns. (9.1) and (9.2) read

$$\mathbf{A}X = (\mathbf{A}_1 \ldots \mathbf{A}_m A_{m+1} \ldots A_n) \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix} = B \qquad . \ (9.1a)$$

and $$f = f(X) = C_t X = a \text{ maximum} \qquad . \quad . \ (9.2a)$$

where X_1 and X_2 are m-dimensional and $(n - m)$-dimensional vectors respectively.

A solution to eqn. (9.1a) which, however, will not in general make f a maximum is

$$X' = \begin{Bmatrix} X_{10} \\ 0 \end{Bmatrix} \qquad . \quad . \quad . \quad . \ (9.3)$$

where $$X_{10} = (A_1 \ldots A_m)^{-1} B \qquad . \quad . \quad . \ (9.4)$$

and the complete solution is

$$X = X' + \text{the } (n - m)\text{-dimensional nullspace of } \mathbf{A}$$

Bearing in mind that the coordinates of X must be non-negative, the solution space (or cell) K of eqn. (9.1a) is the intersection of a non-centred E_{n-m} and the positive n-dimensional orthant (i.e. the 2^nth part of n-space in which all coordinates are $\geqslant 0$).

It now remains to determine the vector X belonging to the cell K which yields a maximum scalar product f with the constant "profit vector" C.

9.2. Definitions

Before proceeding to a detailed discussion of the manner in which the maximum value of f is computed, we shall have to define a number of new concepts.

Definition 1: *Convex Polyhedron* (*or Convex Cell*). Let $V_1, \ldots, V_p$ be a set of vectors; then the set of vectors X such that

$$X = v_1 V_1 + \ldots + v_p V_p \qquad . \quad . \quad . \ (9.5)$$

where $v_i \geqslant 0$, and

$$\sum_{i=1}^{p} v_i = 1$$

is termed the *convex polyhedron* (or *cell*) generated by $V_1, \ldots, V_p$. We shall also use the expression *convex linear combination* for such a combination of vectors.

Exercise. The simplest convex polyhedron consists of a single point defined by one position vector V_1. In this case $p = 1$, and $v_1 = 1$.

When $p = 2$, the polyhedron becomes the line segment connecting the terminal points of the generating vectors V_1 and V_2, and we have

$$\begin{aligned} X &= vV_1 + (1 - v)V_2 \\ &= V_2 + v(V_1 - V_2) \qquad (0 \leqslant v \leqslant 1) \end{aligned}$$

An alternative definition of a convex set is a set of points such that, if X_1 and X_2 are the position vectors of any two points belonging to the set, then all the points on the segment joining them also belong to the set.

Definition 2: *Extreme Point* (*Vertex*) *of a Convex Set.* An extreme point of a convex set is a point which does not lie on a segment joining two other points of the set.

Theorem 37. When the generators of a convex polyhedron are linearly independent, a point of the set will be an extreme point when, and only when, it is a point determined by one generator.

This theorem follows immediately when we recall that none of the generating vectors can be expressed as a linear combination of the others. Hence the equation

$$V_k = v_1 V_1 + \ldots + v_k V_k + \ldots + v_p V_p$$

or $$v_1 V_1 + \ldots + (v_k - 1) V_k + \ldots + v_p V_p = 0$$

has the unique solution $v_k = 1$, and hence $v_i = 0$ $(i \neq k)$, which proves that a generator determines an extreme point.

Conversely, let W be an extreme point; then W must be one of the generators, for, if it were not, it would be a unique linear combination of two or more of the generating vectors, thus contradicting the hypothesis that it is an extreme point. This proves the theorem.

9.3. Properties of the Solution Cell K

We have seen in Section 9.1 that the solutions to eqns. (9.1) comprise a set K which is the intersection of the positive n-dimensional orthant and the non-centred solution space E_{n-m}.

Let V_1 and V_2 be two such solutions; then, from eqn. (9.1a),

$$\begin{aligned} \mathbf{A}(v_1 V_1 + v_2 V_2) &= v_1 \mathbf{A} V_1 + v_2 \mathbf{A} V_2 = v_1 B + v_2 B \\ &= (v_1 + v_2) B = B \qquad . \quad . \quad . \ (9.6) \end{aligned}$$

which means that any point on the convex segment spanned by V_1 and V_2 is also a solution to eqns. (9.1). This result enables us to state the next theorem.

Theorem 38. The solutions of eqns. (9.1) form a convex set. Any point of K can therefore be written as a convex linear combination of the generators K_i ($i = 1, \ldots, p \leqslant n$) of K.

The scalar product of C and X, where

$$X = \sum_1^p k_i K_i \qquad (k_i \geqslant 0 \quad \text{and} \quad \Sigma k_i = 1)$$

is thus

$$f = C_t \Sigma k_i K_i = \Sigma k_i (C_t K_i)$$

Let $C_t K_k = M$ be the largest of the terms $C_t K_i$. Seeing that the k_i are non-negative, it is clear that f will not decrease if we substitute M for $C_t K_i$ in the above expression for f. Hence

$$f = \Sigma k_i (C_t K_i) \leqslant \Sigma k_i M = M$$

If $C_t K_i$ takes on the same maximum value M for a number (say $i = 1, \ldots, q \leqslant p$) of the K_i, then any convex linear combination of these generators will yield the same scalar product with C.

$$C_t(\sum_1^q k_i K_i) = \Sigma k_i (C_t K_i) = \Sigma k_i M = M \qquad . \quad (9.7)$$

These results are summed up in Theorem 39.

Theorem 39. The scalar product $C_t X$, where C is a constant vector and X belongs to the convex set K, takes on its maximum value at an extreme point of K. If it is a maximum at more than one such point, any convex linear combination of these points will yield the same maximum value.

Yet another important property of the extreme points of K is given in the following theorem.

Theorem 40. The number of non-zero coordinates of an extreme point X of the convex cell K never exceeds m, and the m-dimensional vectors A_i (*see* eqn. (9.1*a*)) corresponding to the non-zero coordinates of X are linearly independent.

To prove this theorem, suppose $p > m$ coordinates of X are positive, and let X be an extreme point of K. Without any loss of generality we can assume that the first p coordinates of X are > 0. We now have

$$x_1 A_1 + \ldots + x_p A_p = B \qquad . \quad . \quad . \quad (9.8)$$

where the p m-dimensional vectors A_i must be linearly dependent because $p > m$. This means that

$$\Sigma d_i A_i = 0 \qquad . \quad . \quad . \quad . \quad (9.9)$$

where not all the d_i's are zero.

Multiplying eqn. (9.9) by $\pm h$ and adding it to eqn. (9.8), we get

$$\sum_1^p (x_i \pm h d_i) A_i = B$$

By choosing h sufficiently small we can make all the coefficients $x_i \pm h d_i$ positive ($x_i > 0$), which in turn means that the n-dimensional vectors X_1 and X_2 with coordinates $x_i + h d_i$ and $x_i - h d_i$, respectively, both satisfy eqn. (9.8) and generate a convex segment which contains X ($= \frac{1}{2}X_1 + \frac{1}{2}X_2$), thus contradicting our hypothesis that X is an extreme point. This proves that the A_i's ($i = 1, \ldots, p$) are independent, and therefore $p \leqslant m$.

The solution to the L.P. problem must thus be sought at the intersection of the solution E_{n-m} (which is non-centred) and the positive coordinate E_m's of the n-dimensional space in which the problem is immersed. We have shown in Exercise 1, Section 4.5, that the general intersection in an E_n of an E_{n-m} and an E_m is an E_0 (a point).

9.4. Dantzig's Simplex Method

Up to now we have stated the problem of L.P. and established a number of theorems concerning the properties of the solution cell K and the required vector X. In this section we shall sketch briefly a technique for determining the optimal solution to the problem.

First we shall explain the purpose and use of *slack variables* and *artificial unit vectors*.

Slack variables are introduced in order to change an inequality into an equation. Suppose one of the members of eqn. (9.1) had read

$$a_1 x_1 + a_2 x_2 \leqslant b \qquad . \quad . \quad . \quad (9.10)$$

To satisfy this inequality the components of X in the x_1-x_2 plane must lie in the shaded area of Fig. 9.1(*a*).

OAB is a convex cell which cannot be generated in 2-space by a set of linearly independent vectors. By adding an extra dimension to our space, as shown in Fig. 9.1(*b*), we avoid this difficulty and inequality (9.10) becomes

$$a_1 x_1 + a_2 x_2 + x_{n+1} = b \qquad . \quad . \quad . \quad (9.11)$$

where $x_{n+1} \geqslant 0$.

The extra coordinate x_{n+1} does not represent any actual product, and the corresponding coordinate of the augmented vector C is zero. If the final solution contains non-zero slack variables, this will be an indication that maximum profit can be achieved when certain machines in the plant are running at less than full capacity. Each slack variable x_i ($i = n + 1, \ldots$) will be associated with an m-dimensional unit vector U_i in the matrix of constraint **A**.

The introduction of artificial unit vectors is a device designed to evade the necessity of inverting the matrix given in eqn. (9.1*a*). Originally, matrix **A** may or may not contain m-dimensional unit

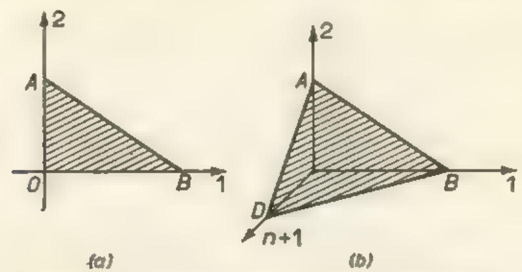


FIG. 9.1. INTRODUCTION OF SLACK VARIABLES

column vectors. Every slack variable has added a unit vector to the set. We now complete the set of unit vectors by adding artificial unit vectors to make up a total of m.

The associated variables x_i do not correspond to any products, and to ensure that they shall not appear in the final solution, we make the corresponding coordinates of C equal to $-P$, where P is an unspecified large positive number.

After renumbering the coordinates of X so as to place the m unit vectors U_i in the first m columns of the augmented **A** matrix, we arrive at the equation

$$x_1U_1 + \ldots + x_mU_m + x_{m+1}A_{m+1} + \ldots + x_nA_n = B \quad . \ (9.12)$$

Each vector of the left-hand member of this equation is associated with a coordinate of the profit vector C. In the case of a vector which was present in the orginal matrix of constraint (**A**), the corresponding coordinate is c_i, while the unit vectors that have been introduced in connexion with the slack and artificial variables correspond to coordinates of C that are 0 and $-P$, respectively.

By inspection it is seen that

$$x_i = \begin{cases} b_i & \text{for } i = 1, \ldots, m \\ 0 & \text{for } i = m + 1, \ldots n \end{cases}$$

is a solution of equation (9.12) which must represent an extreme point of K, because the column vectors of **A** associated with the non-zero coordinates of X are independent.

The total profit for this X is

$$f = C_tX = \sum_{i=1}^{m} c_ib_i \quad . \quad . \quad . \ (9.13)$$

which, however, is not as a rule the required maximum.

Using this vertex of K as our point of departure, we now apply the simplex method by substituting a vector from the set

$$A_{m+1}, \ldots, A_n$$

for one of the m unit vectors U_i in such a way that—

(i) Eqn. (9.12) is still satisfied.
(ii) f is increased as much as possible.
(iii) None of the new x_i's is allowed to become negative.

The first m vectors $U_1, \ldots, U_m$ of **A** form a base in terms of which all the other column vectors of **A** can be expressed. Thus

$$A_k = a_{1k}U_1 + \ldots + a_{mk}U_m \quad . \quad . \ (9.14)$$

where a_{ik} are the m coordinates of A_k.

Multiplying eqn. (9.14) by h and adding it to eqn. (9.12), we find that

$$\sum_1^m (b_i - ha_{ik})U_i + hA_k = B \quad . \quad . \ (9.15)$$

The coordinates of the new vector X' are

$$x_i' = \begin{cases} b_i - ha_{ik} & \text{for } i \leqslant m \\ h & \text{for } i = k > m \\ 0 & \text{for } (i \neq k) > m \end{cases}$$

and the new value of f is

$$\begin{aligned} f' &= \sum_1^m c_i(b_i - ha_{ik}) + hc_k \\ &= f + h(c_k - \sum_1^m c_ia_{ik}) \\ &= f + h(c_k - g_k) \quad . \quad . \quad . \quad . \ (9.16) \end{aligned}$$

Eqn. (9.16) enables us to pick the value of k that will result in the greatest increase,

$$f' - f = h(c_k - g_k)$$

in the total profit; h is positive (it is a coordinate of X'), and we shall therefore choose k such that

$$c_k - g_k \,(> 0) = \text{a maximum}$$

This done, the choice of h is governed by the consideration that one of the

$$x_i = b_i - ha_{ik} \qquad (i \leqslant m)$$

must be zero and the remainder greater than zero.

Hence $\qquad h = (b_i/a_{ik})_{\min} \qquad (i = 1, \ldots, m)$

where only positive a_{ik}'s are considered.

In this manner we obtain a maximum increase in f, and at the same time elminate one of the terms x_i, while keeping the remainder non-negative.

Let the rth coordinate of $X\,(r \leqslant m)$ be the one that vanishes. Eqn. (9.15) then becomes

$$\left(b_1 - \frac{b_r}{a_{rk}} a_{1k}\right) U_1 + \ldots + 0\,.\,U_r + \ldots + \left(b_m - \frac{b_r}{a_{rk}} a_{mk}\right) U_m + \frac{b_r}{a_{rk}} A_k = B \qquad (9.17)$$

The coordinates of the new position vector X' can be read from eqn. (9.17). It will be noted that the rth coordinate vanishes and that coordinate k, which was zero in X, now becomes $x_k' = h = b_r/a_{rk}$.

Another way of interpreting the coefficients of the vectors in eqn. (9.17) is to look upon them as the coordinates of B in the coordinate system formed by the vectors A_k and $U_i\,(i = 1, \ldots, r-1, r+1, \ldots, m)$.

If we wish to calculate the coordinates of any other vector (say A_j) referred to this base, we proceed as follows. Solving eqn. (9.14) for U_r, we get

$$U_r = \frac{1}{a_{rk}}\left(A_k - \sum_{i \neq r}^{m} a_{ik} U_i\right) \qquad (9.18)$$

After making $k = j$ in eqn. (9.14), and employing the above expression for U_r, we obtain an expression for A in terms of the new base—

$$A_j = \sum_{i \neq r}^{m} \left(a_{ij} - \frac{a_{rj}}{a_{rk}} a_{ik}\right) U_i + \frac{a_{rj}}{a_{rk}} A_k \qquad (9.19)$$

where $m < j \leqslant n$.

Using these values for the coordinates of B and the vectors A_j, we prepare for the next step of the simplex analysis by compiling

the new values of $c_i - g_i$ in order to decide whether we have reached the final solution (all $c_i - g_i = 0$), or whether it is still possible to increase f by replacing one of the vectors of the base by a new A_k.

9.5. Solutions at Infinity—Multiple Maxima—Degeneracy

An exhaustive discussion of the simplex method lies outside the scope of this book. We shall therefore have to content ourselves by giving a geometrical explanation of the various difficulties that may arise during its application.

To recapitulate: the problem of linear programming is to determine a vector X, constrained to lie within an $(n - m)$-dimensional cell K in n-space, such that the projection of X onto a vector C is as large as possible.

We have proved that K is a convex cell, that the point or points of K which make f a maximum form a convex subset of K (usually consisting of one vertex), and finally that the position vectors of the vertices of K never have more than m non-zero coordinates.

In the case $n = 3$ we are able to visualize the problem directly. When $m = 1$, the convex solution cell K is a triangle spanned by three points on the positive axes; whereas, when $m = 2$, it consists of a line segment connecting two points in the positive coordinate planes.

Both these cases are illustrated in Fig. 9.2.

In the higher-dimensional case our spatial intuition fails us, but we can still picture K as consisting of a number of vertices connected by line segments (edges) to form what topologists call a connected graph.

Geometrically speaking, the simplex method consists in moving along the edges of the solution cell from vertex to vertex in such a way that each step increases the value of f.

The numerical calculations can be arranged in a *tableau* so as to show at a glance which adjacent vertex will give the greatest increase in f.

It is necessary to know the coordinates of one vertex (any one) of K before the simplex method can be employed. The introduction of artificial variables is, in fact, a technique by which the basic n-space of the analysis is augmented by the addition of extra reference axes. By this means a vertex is created from which to start the search for the maximizing point of K. The severe penalty $-P$ attached to the artificial variables ensures that they cannot appear in the final solution; in other words, we are eventually forced out of the artificial annex and back into our basic n-space.

Three special situations may arise during the computation—

(i) It was shown in Section 9.3 that the vertices of K must be sought at the intersections of the non-centred solution E_{n-m} of the constraining eqn. (9.1*a*) and the positive coordinate E_m's of the n-dimensional coordinate system.

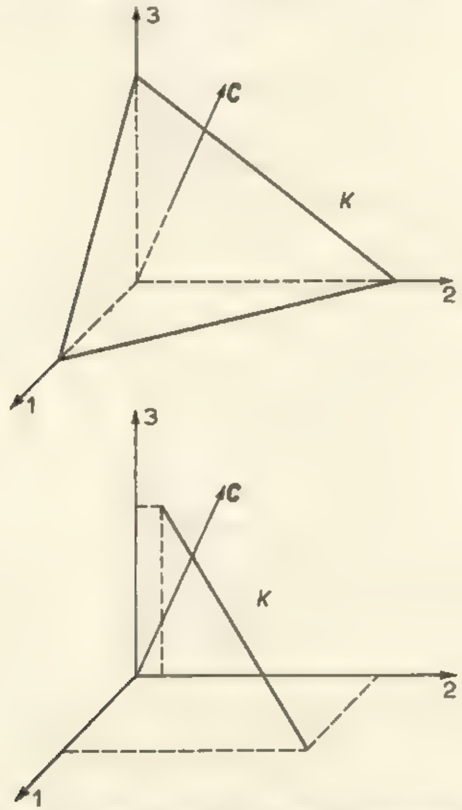


Fig. 9.2. The Problem of Linear Programming

If, however, the E_{n-m}'s and a coordinate E_m happen to have a p-direction in common, the corresponding vertices lie at infinity; also, if one (or more) of the m non-zero coordinates of a vertex is negative, the cell K will be unbounded. In either case the L.P. analysis may lead to the result that the total profit f will increase without limit when certain coordinates of X tend to infinity.

(ii) A second type of difficulty occurs when a convex subcell K_1 of K is at right angles to C. All the position vectors of K_1 will

have the same scalar product with C, and if the corresponding value, f_1, is a maximum, the manufacturer can employ any vector in K_1 as the basis of his production schedule. His final choice of X will now have to be decided by considerations not originally incorporated in eqn. (9.1*a*).

Under these circumstance Theorem 40, p. 102, can be ignored and the solution may comprise more than m non-zero x_i's.

(iii) Finally, there exists a type of computational irregularity that has been given the name *degeneracy*. It occurs when more than one unit vector disappears during the process of changing the base vectors A_i (*see* eqn. (9.17)). The resulting vector X has less than m positive coordinates. This difficulty, which is inherent in the simplex method, causes uncertainty as to the next step to be taken in the vertex-by-vertex search for a maximal solution. It can be overcome by means of a perturbation technique equivalent, geometrically, to a slight linear distortion of the solution cell K.

PROBLEMS

9.1. In connexion with Theorem 37, show that when the generators of K are not linearly independent, an extreme point is determined by a generator, but that the converse statement need not be true.

9.2. Determine a 4-dimensional vector X lying in the positive orthant and subject to the constraining equations

$$\begin{Bmatrix} 0 & 0 & 2 & 1 \\ -7 & 1 & 2 & -5 \\ 6 & 0 & -4 & 1 \end{Bmatrix} X = \begin{Bmatrix} 10 \\ 14 \\ 4 \end{Bmatrix}$$

such that its scalar product with the profit vector C with coordinates (2, −1, 10, 20) is a maximum.

CHAPTER 10

Linear Network Analysis

10.1. Preliminaries and Definitions

LINEAR network analysis is fundamentally important to an electrical engineer. In this chapter we shall show how matrix algebra can be applied with advantage to the study of linear networks, and how geometry and topology combine to shed quite a new light on the subject.

The adjective "linear" can be applied to a network when the elements from which it is built (resistors, capacitors, self- and mutual inductors, etc.) are proper constants that are unaffected by the currents flowing through them or the voltages applied across them. This condition is very nearly satisfied for a great variety of the networks in practice. The differential equations describing the performance of a linear network have constant coefficients, and the condition of linearity (*see* eqn. (6.11)) is therefore satisfied: When certain excitations $E_1(t)$ and $E_2(t)$ of a network result in responses $R_1(t)$ and $R_2(t)$, respectively, the combined excitation $k_1E_1(t) + k_2E_2(t)$ will cause the response $k_1R_1(t) + k_2R_2(t)$ for any values of k_1 and k_2. In practice, of course, k_1 and k_2 will be allowed to vary only within specified limits.

Topologically, any network is described in terms of the following concepts—

(i) *Branches*. A branch consists of one element or a number of elements in series; all elements of a branch carry the same current under all conditions.

(ii) *Nodes*. The terminals by means of which a branch can be connected to other branches are called nodes (junctions).

(iii) *Node-pairs*. Any two nodes of a subnetwork (*see* (vi)) constitute a node-pair. A connected network with n nodes has $\frac{1}{2}n(n-1)$ node-pairs.

(iv) *Meshes*. A closed path traced through a network is called a mesh (loop).

(v) *Open Meshes*. An open path traced through a network, starting at one node and ending at another, is termed an open mesh.

(vi) *Subnetworks*. A connected subset of branches not electrically connected with the rest of the network is called a subnetwork. Subnetworks may, however, be coupled magnetically. Also, two networks with only one node in common can be considered to be distinct subnetworks.

(vii) *Graphs*. The graph of a network is a diagram in which each branch of the network is indicated by an oriented line.

(viii) *Trees*. A tree on a connected network is a graph corresponding to a set of branches connecting all the modes of the network without forming any meshes.

(ix) *Cotrees*. A cotree on a connected network is the graph of the complement of a tree on the network (i.e. the graph of the branches that remain when the branches of the tree have been removed). The branches of a cotree are called *links*.

(x) *Cut-sets*. A cut-set (a set of tie branches) of a network is a minimal set of branches which, when detached, divides the network into at least two subnetworks.

(xi) *Cocut-sets*. A cocut set (a set of coties) is a minimal set of branches which, when short-circuited, divides the network into at least two subnetworks (which will have a node (or nodes) in common).

With the help of the concepts *tree* and *cotree*, we can prove a fundamental topological theorem known as Euler's formula.

Let us construct a network by first joining branches together to form a covering tree, and then adding branches of the corresponding cotree until the network is complete.

The first branch of a connected network contributes two nodes, and for each subsequent branch that we add to the tree, we also add a new node. If the network comprises s unconnected subnetworks, we thus find

$$b_t = n - s \stackrel{D}{=} p \quad . \quad . \quad . \quad . \quad (10.1)$$

where b_t is the total number of branches in a tree on the network, and n the number of nodes.

When we connect a branch belonging to the cotree, we are not creating any more nodes, since by definition the branches of the tree connect all the nodes. Each new branch will, however, complete a mesh. This is clear, since a branch of the cotree must join two nodes of the tree, and these nodes are connected through the tree via a unique path. Such a mesh passing through the tree plus one branch of the cotree (a link), is termed a *basic mesh*. Thus we conclude that

$$b_{t'} = m \quad . \quad . \quad . \quad . \quad (10.2)$$

where $b_{t'}$ is the total number of branches (links) in the cotree. By adding eqns. (10.1) and (10.2), we get

$$b_t + b_{t'} = b = m + (n - s) = m + p \qquad . \quad (10.3)$$

which, expressed in words, reads—

Theorem 41 (*Euler's Formula*). The number of branches (b) of a network is equal to the number of independent meshes (m) plus the number of independent node-pairs ($n - s = p$).

This theorem does not depend on whether or not the network is planar.

Problems in network analysis are usually stated as follows. The b branch impedors together with the mutual couplings between them are known. Also, certain impedanceless sources of constant e.m.f. in series with the branches and certain admittanceless sources across

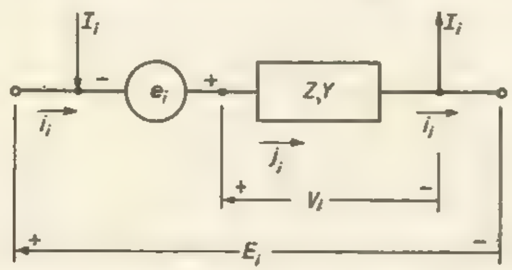


FIG. 10.1. POLARITY CONVENTIONS

the branches are given. It is required to determine the branch currents and the voltages appearing across the branches. This is done by means of Kirchhoff's two equations in conjunction with Ohm's law.

Kirchhoff's first law (Kirchhoff I) states that, at any instant, the sum of all the currents flowing towards a node is equal to the sum of all the currents flowing away from it. This self-evident idea implies the conservation of electric charge and has its generalized counterparts (using the divergence operator) in aero- and thermodynamics. Kirchhoff I yields p independent constraining relations between the branch currents.

According to Kirchhoff's second law (Kirchhoff II), the sum at any instant of the potential differences (active and passive) round a closed mesh is zero. This is a reformulation of the principle of the conservation of energy, and leads to m independent constraining

equations between the branch p.d.s. The generalized version of Kirchhoff II uses the rotation (or curl) operator.

To avoid confusion and errors, it is imperative that a definite system of orientation of currents and voltages be decided upon and strictly adhered to. For the sake of generality, we shall assume that each branch contains a series-connected source of e.m.f. e_i and a parallel-connected source of current I_i, as indicated in Fig. 10.1.

The meaning of the symbols is obvious in the case of the currents, and the polarity convention for voltages is such that the p.d. between two nodes is taken as positive when the node at the arrow-head is at a higher potential than the other.

Hence, the current flowing through the branch *immittance* (impedance or admittance) is

$$j_i = i_i + I_i \qquad (i = 1, \ldots, b) \qquad . \quad . \quad (10.4)$$

and the p.d. appearing across it is

$$V_i = e_i + E_i \quad . \quad . \quad . \quad . \quad (10.5)$$

Exercise. In order to illustrate the concepts introduced in this section, let us consider the network shown in Fig. 10.2. The circuit consists of 2 ($= s$) magnetically coupled subnetworks N_1 and N_2. It comprises 8 ($= n$) nodes ($A, B, \ldots, H$), and 13 ($= b$) branches ($a, b, \ldots, o$). Any two nodes belonging to the same subnetwork (CE, AB, HG, etc.) form a node-pair. Subnetwork N_1,

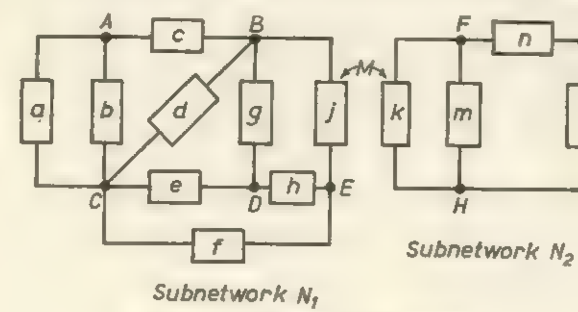


FIG. 10.2. EXAMPLE OF NETWORK

for instance, comprises $\frac{1}{2}n_1(n_1 - 1) = \frac{1}{2} \times 5 \times 4 = 10$ node-pairs, the p.d.s of which can be expressed as linear combinations of $n_1 - 1 = 4$ independent node-pair p.d.s.

A closed path such as $ABDECA$ through the network is a mesh, and by means of Euler's theorem we can calculate the number of independent meshes to be $m = b - n + s = 13 - 8 + 2 = 7$.

The branches b, d, g and h form a tree on subnetwork N_1, and the remaining branches a, c, e, j and f make up the corresponding cotree. It is readily seen that

a number of other trees exist on N_1, and also that any such tree must contain exactly $n_1 - 1 = 4$ branches. The reader should choose a tree on N_1 and then indicate the corresponding basic meshes, each of which links with a branch of the cotree and is completed by means of a unique path through the tree.

If the branches c, d, e and f are detached from N_1, the subnetwork will be divided into two unconnected parts. These branches therefore form a cut-set of N_1. Branch d forms a cocut-set of N_1, since, when d is short-circuited, N_1 will be separated into two networks which have only one node ($B = C$) in common.

10.2. Connexion Matrices—Kirchhoff's Laws

It was mentioned in the previous section that each branch has a current i_t flowing through it. We shall arrange these currents as a b-dimensional vector and denote it by a single symbol $\boldsymbol{i}$. The coordinates of this vector are not independent, however. The branch currents must obey Kirchhoff I, which allows $\boldsymbol{i}$ only m degrees of freedom. Maxwell realized this fact when he introduced his m independent mesh currents $\boldsymbol{i}'$. The relation between $\boldsymbol{i}$ and $\boldsymbol{i}'$ can be expressed by means of a b-by-m *mesh-connexion matrix* $\mathbf{C}$.

$$\boldsymbol{i} = \mathbf{C}\boldsymbol{i}' \quad . \quad . \quad . \quad . \quad (10.6)$$

The connexion matrix uniquely determines the way in which the branches are connected to form a network (*see* Exercise 2 below). The converse is not true.

With the introduction of independent mesh currents, Kirchhoff I is automatically satisfied. To derive an expression for the driving e.m.f.s acting in the m meshes, we make use of the fact that the instantaneous input of power to the network cannot depend on whether we consider branch or mesh quantities. Thus

$$P = \boldsymbol{i}_t\boldsymbol{e} = \boldsymbol{i}'_t\boldsymbol{e}' \quad . \quad . \quad . \quad . \quad (10.7)$$

where $\boldsymbol{e}$ is the b-dimensional column vector containing all the branch e.m.f.s and $\boldsymbol{e}'$ is an m-dimensional vector whose coordinates are the mesh e.m.f.s.

Substituting eqn. (10.6) in eqn. (10.7), we get

$$\boldsymbol{i}'_t\mathbf{C}_t\boldsymbol{e} = \boldsymbol{i}'_t\boldsymbol{e}' \quad . \quad . \quad . \quad . \quad (10.8)$$

from which we can conclude that

$$\boldsymbol{e}' = \mathbf{C}_t\boldsymbol{e} \quad . \quad . \quad . \quad . \quad (10.9)$$

since eqn. (10.8) must be valid for all values of $\boldsymbol{i}'$.

Using the branch p.d.s E_t as our unknowns, we can develop the dual of the mesh method of analysis when we note that the E_t's

are not independent but must meet the requirements of Kirchhoff II. Expressing the b E_t's in terms of p independent node-pair voltages $\boldsymbol{E}'$, we arrive at the dual of eqn. (10.6)—

$$\boldsymbol{E} = \mathbf{A}\boldsymbol{E}' \quad . \quad . \quad . \quad . \quad (10.10)$$

where $\mathbf{A}$ is called the *node-pair connexion matrix*. By analogy with the term *basic mesh*, defined on p. 111, we term a node-pair across a branch of a tree on a network a *basic node-pair*.

The currents $\boldsymbol{I}$ impressed across the branches can be expressed as linear combinations of a set of currents $\boldsymbol{I}'$ injected across the p independent node-pairs—

$$\boldsymbol{I}' = \mathbf{A}_t\boldsymbol{I} \quad . \quad . \quad . \quad . \quad (10.11)$$

the proof of this formula being the dual of that of eqn. (10.9). In this case, $P = \boldsymbol{I}_t\boldsymbol{E} = \boldsymbol{I}'_t\mathbf{A}_t\boldsymbol{E} = \boldsymbol{I}'_t\boldsymbol{E}'$ is an invariant.

A fascinating relationship exists between the two connexion matrices of a network. Both have b rows, so that the pre-product of the one by the transpose of the other is always defined.

Viewed as linear operators their significance when employed as pre-factors is set out in Table 10.1.

Table 10.1

OPERATIONS PERFORMED BY CONNEXION MATRICES

Matrix	Shape	Operation
$\mathbf{C}$	b-by-m	combines m mesh quantities to form b branch quantities
$\mathbf{C}_t$	m-by-b	sums branch quantities acting round m meshes
$\mathbf{A}$	b-by-p	combines p node-pair quantities to form b branch quantities
$\mathbf{A}_t$	p-by-b	sums the branch quantities converging on p nodes

Suppose m independent mesh currents $\boldsymbol{i}'$ are given; the branch currents are then $\boldsymbol{i} = \mathbf{C}\boldsymbol{i}'$. Kirchhoff I is automatically satisfied, so that, when we sum the branch currents flowing towards the p independent nodes of the network, the result is a p-dimensional null-vector. Thus

$$\mathbf{A}_t\mathbf{C}\boldsymbol{i}' = \boldsymbol{0} \quad . \quad . \quad . \quad . \quad (10.12)$$

This identity holds for all $\boldsymbol{i}'$, so therefore

$$\mathbf{A}_t\mathbf{C} = \mathbf{0} \quad . \quad . \quad . \quad . \quad (10.13)$$

Apparently, the column vectors of $\mathbf{C}$ span the nullspace of $\mathbf{A}_t$ and vice versa.

Exercise 1. Compute **C** and **A** for the bridge network of Fig. 10.3, and test the validity of eqn. (10.13).

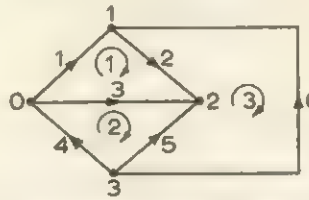


FIG. 10.3. TOPOLOGICAL GRAPH OF BRIDGE NETWORK

In this case $b = 6, n = 4, s = 1, p = 3$ and $m = 3$. After numbering and orientating the branches, node-pairs and meshes, we can compute the connexion matrices—

$$\mathbf{C} = \begin{Bmatrix} 1 & 0 & 0 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{Bmatrix}$$

and

$$\mathbf{A} = \begin{Bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{Bmatrix}$$

It is readily verified that $\mathbf{A}_t\mathbf{C} = \mathbf{0}$, and, by transposition, that $\mathbf{C}_t\mathbf{A} = \mathbf{0}$.

Exercise 2. Reconstruct the network from either the mesh connexion matrix **C** or the node-pair connexion matrix **A**.

We shall not discuss the solution of this problem in detail, but merely point out that each column of **C** indicates and orients the branches forming the corresponding mesh. Similarly, each column of **A** lists and orients the branches converging on the corresponding node. The reader will discover for himself that it is usually easier to construct the network from matrix **A** than from **C**.

10.3. Mesh Method of Analysis

Irrespective of the configuration of the network, Ohm's law must apply to each and every branch immittance: The potential difference V_i will include, not only the voltage drops caused by the branch currents, but also, in the case where magnetic coupling exists between the branches, e.m.f.s induced in the branch immittance by currents in other branches.

If we arrange the self- and mutual inductances of the branches in a b-by-b impedance matrix **Z** (the *primitive* impedance matrix), we can write Ohm's law simultaneously for all branches as

$$V = E + e = \mathbf{Z}(i + I) = \mathbf{Z}J \quad . \quad . \quad (10.14)$$

The elements along the main diagonal of **Z** are the self-impedances of the branches, and the element in position ij is the transimpedance of branch j into branch i.

Summing the V round the m meshes is accomplished, as shown in Table 10.1, by pre-multiplication by $\mathbf{C}_t$. Hence

$$\mathbf{C}_tV = \mathbf{C}_tE + \mathbf{C}_te = 0 + e' = \mathbf{C}_t\mathbf{Z}\mathbf{C}i' + \mathbf{C}_t\mathbf{Z}I \quad . \quad (10.15)$$

where $\mathbf{Z}' = \mathbf{C}_t\mathbf{Z}\mathbf{C}$ is the m-by-m *mesh impedance matrix* of the given network.

Solving eqn. (10.15) for i' and using eqns. (10.6) and (10.9), we arrive at an expression for the required branch currents in terms of the applied branch e.m.f.s e, the injected branch currents I, the primitive impedance matrix **Z**, and the connexion matrix **C**—

$$i = \mathbf{C}(\mathbf{C}_t\mathbf{Z}\mathbf{C})^{-1}\mathbf{C}_t(e - \mathbf{Z}I) \quad . \quad . \quad (10.16)$$

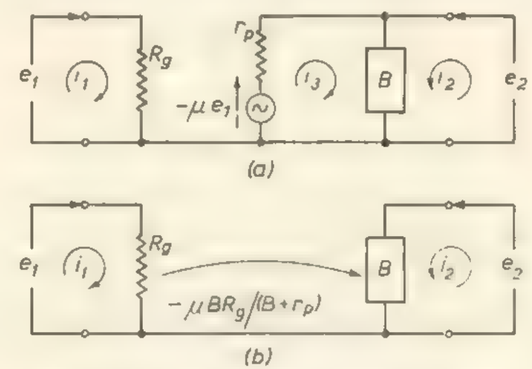


FIG. 10.4. EQUIVALENT MESH CIRCUIT OF TRIODE

When the only active sources in the network are the series branch e.m.f.s e (i.e. when $I = 0$), eqn. (10.16) becomes

$$i = \mathbf{C}(\mathbf{C}_t\mathbf{Z}\mathbf{C})^{-1}\mathbf{C}_te \quad . \quad . \quad . \quad (10.17)$$

To obtain the unknown node-pair p.d.s, we solve eqn. (10.14) for E and substitute expression (10.16) for i. Hence, after some manipulation,

$$E = (\mathbf{Z}\mathbf{C}(\mathbf{C}_t\mathbf{Z}\mathbf{C})^{-1}\mathbf{C}_t - \mathbf{I})(e - \mathbf{Z}I) \quad . \quad . \quad (10.18)$$

In the case of passive networks, the impedance matrix **Z** is symmetric. When, however, an active element, such as a vacuum tube, is connected in the circuit, the matrix immediately loses its symmetry, as we shall demonstrate by the following example.

Consider the circuit of Fig. 10.4. For Class A operation the triode can be replaced by its equivalent circuit as shown at (*a*), and we arrive at the three mesh equations,

$$\begin{aligned} e_1 &= R_g i_1 \\ \mu e_1 &= \qquad -Bi_2 - (B + r_p)i_3 \\ e_2 &= \qquad Bi_2 + \qquad Bi_3 \end{aligned}$$

Eliminating i_3 and writing the equations in matrix form, we find

$$e = \begin{Bmatrix} e_1 \\ e_2 \end{Bmatrix} = \begin{Bmatrix} R_g & 0 \\ -\mu BR_g/(B + r_p) & Br_p/(B + r_p) \end{Bmatrix} \begin{Bmatrix} i_1 \\ i_2 \end{Bmatrix} = \mathbf{Z}i$$

Viewed as a 2-branch 2-mesh circuit (Fig. 10.4(*b*)), a network containing a vacuum tube is represented by a non-symmetric impedance matrix $\mathbf{Z}$.

It is possible to resolve $\mathbf{Z}$ into two parts. One of them,

$$\mathbf{Z}_n = \begin{Bmatrix} R_g & 0 \\ 0 & B \end{Bmatrix}$$

is the branch matrix of the network before connexion of the tube, and the other,

$$\mathbf{Z}_v = \begin{Bmatrix} 0 & 0 \\ -\mu BR_g/(B + r_p) & -B^2/(B + r_p) \end{Bmatrix}$$

which is unsymmetric, is due to the action of the tube. This decomposition is not particularly satisfactory, however, as the elements of $\mathbf{Z}_v$ are also functions of the grid and load resistances, so that $\mathbf{Z}_v$ is not truly a tube matrix. We shall return to this question in the next section.

10.4. Node-Pair Method of Analysis

With eqn. (10.14) as our starting-point, we obtain the expression

$$J = I + i = \mathbf{Y}(E + e) = \mathbf{Y}V \qquad . \qquad . \quad (10.19)$$

where $\mathbf{Y} = \mathbf{Z}^{-1}$ is termed the *primitive* admittance matrix of the network.

By following a train of reasoning which is the exact dual of that employed in Section 10.3, we derive a formula expressing the unknown node-pair voltages in terms of the injected currents I, the series-connected branch e.m.f.s e, the primitive admittance matrix $\mathbf{Y}$, and the node-pair connexion matrix $\mathbf{A}$—

$$E = \mathbf{A}(\mathbf{A}_t\mathbf{Y}\mathbf{A})^{-1}\mathbf{A}_t(I - \mathbf{Y}e) \qquad . \qquad . \quad (10.20)$$

This equation is the dual of eqn. (10.16).

Similarly,

$$i = (\mathbf{Y}\mathbf{A}(\mathbf{A}_t\mathbf{Y}\mathbf{A})^{-1}\mathbf{A}_t - \mathbf{I})(I - \mathbf{Y}e) \qquad . \qquad . \quad (10.21)$$

is the dual of eqn. (10.18).

When $e = 0$ (i.e. when the excitation consists of injected currents I only), eqn. (10.20) reduces to the form

$$E = \mathbf{A}(\mathbf{A}_t\mathbf{Y}\mathbf{A})^{-1}\mathbf{A}_t I \qquad . \qquad . \qquad . \quad (10.22)$$

By analogy with the nomenclature introduced in the previous section, the p-by-p matrix $\mathbf{Y}' = \mathbf{A}_t\mathbf{Y}\mathbf{A}$ is termed the *node-pair admittance matrix*.

In the general case, the currents flowing into the electrodes of a vacuum tube are functions of all the electrode potentials relative to the cathode. This relationship can be most conveniently expressed by means of a square differential admittance matrix $\mathbf{Y}_v$.

For a triode we have

$$\mathbf{Y}_v = \begin{Bmatrix} \partial I_g/\partial E_g & \partial I_g/\partial E_p \\ \partial I_p/\partial E_g & \partial I_p/\partial E_p \end{Bmatrix} \qquad . \qquad . \qquad . \quad (10.23)$$

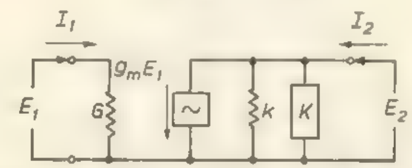


FIG. 10.5. EQUIVALENT NODE-PAIR CIRCUIT OF TRIODE

Changes in the electrode currents I for small changes in the electrode potentials E can be written concisely as

$$dI = \mathbf{Y}_v dE \qquad . \qquad . \qquad . \qquad . \quad (10.24)$$

If the grid of the triode is not allowed to be positive, the elements $\partial I_g/\partial E_p$ and $\partial I_g/\partial E_g$ both vanish and the tube matrix becomes

$$\mathbf{Y}_v = \begin{Bmatrix} 0 & 0 \\ g_m & k \end{Bmatrix} \qquad . \qquad . \qquad . \quad (10.25)$$

where g_m is the grid–anode transconductance, and $k = 1/r_p$ is the anode conductance.

The equivalent admittance circuit of the triode is shown in Fig. 10.5, in which G is the grid-leak conductance and K the load conductance.

The performance of the network is given by the following matrix equation—

$$I = \begin{Bmatrix} G & 0 \\ g_m & K + k \end{Bmatrix} E = \mathbf{Y}E \qquad . \qquad . \quad (10.26)$$

Y resolves quite naturally into two component matrices,

$$\mathbf{Y}_n = \begin{Bmatrix} G & 0 \\ 0 & K \end{Bmatrix}$$

corresponding to the network before the valve was connected, and $\mathbf{Y}_v$ as given in eqn. (10.25), which represents the effect of the valve itself. Its components are functions of the valve constants only, and do not depend on the circuit in which it is connected. Valves are typical node-pair devices and lend themselves more readily to treatment by the node-pair method than by the mesh method.

The tube matrix given in eqn. (10.23) can easily be generalized to cater for multi-electrode tubes.

Exercise. Matrix analysis of a network by, say, the node-pair method results in an expression

$$I' = \mathbf{Y}'E' = \mathbf{A}_t\mathbf{Y}\mathbf{A}E' \quad . \quad . \quad . \quad . \quad (10.27)$$

relating the impressed currents I' to the response p.d.s E'.

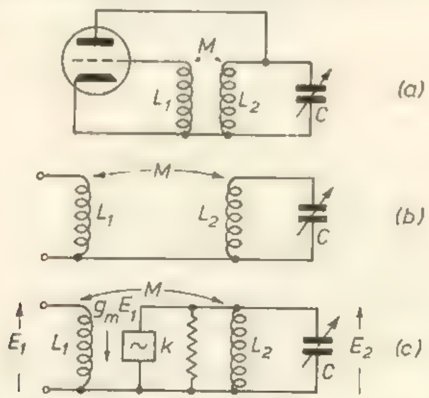


FIG. 10.6. DERIVATION OF VAN DER BIJL'S EQUATIONS

When the node-pair admittance matrix $\mathbf{Y}'$ is invertible there is a unique correspondence between the vectors I' and E'. A given set of injected currents I' will cause certain p.d.s E' to appear across the node-pairs, and when $I' = 0$, E' must also vanish.

If, however, $\mathbf{Y}'$ is singular, this one-to-one relationship no longer exists, and E' may be $\neq 0$ even though $I' = 0$. Under these conditions the circuit is unstable, and voltages appear across the terminals of the network in spite of the fact that no external excitation is applied.

Fig. 10.6(*a*) shows the a.c. circuit of a simple tuned-anode oscillator. When the regenerative coupling between the anode and grid coils is sufficiently tight, the circuit will break into self-oscillation. In his book *The Thermionic Vacuum Tube and its Applications*, van der Bijl derives the criterion for oscillation in an oscillator of this type by studying the differential equations of performance of the circuit. We shall now arrive at the same expression by matrix analysis.

Fig. 10.6(*b*) shows the network before the triode is connected. From it we derive the primitive branch impedance matrix **Z**—

$$Z = \begin{Bmatrix} j\omega L_1 & -j\omega M & 0 \\ -j\omega M & R + j\omega L_2 & 0 \\ 0 & 0 & 1/j\omega C \end{Bmatrix} \quad . \quad . \quad . \quad (10.28)$$

where the mutual impedance between branches 1 and 2 is taken to be negative so as to ensure the correct polarity of feedback.

Y, the primitive admittance matrix, is the reciprocal of **Z**—

$$\mathbf{Y} = \mathbf{Z}^{-1} = \begin{Bmatrix} (R + j\omega L_2)/D & j\omega M/D & 0 \\ j\omega M/D & j\omega L_1/D & 0 \\ 0 & 0 & j\omega C \end{Bmatrix} \quad . \quad . \quad (10.29)$$

where $D = (M^2 - L_1L_2)\omega^2 + j\omega L_1R$ is the subdeterminant derived from **Z** by deleting its last row and last column. $D \neq 0$ when $\omega \neq 0$.

The node-pair connexion matrix of the network is

$$\mathbf{A} = \begin{Bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{Bmatrix}$$

and by means of eqn. (10.27) we find the node-pair connexion matrix of the network to be

$$\mathbf{Y}'_n = \mathbf{A}_t\mathbf{Y}\mathbf{A} = \begin{Bmatrix} (R + j\omega L_2)/D & j\omega M/D \\ j\omega M/D & j\omega L_1/D + j\omega C \end{Bmatrix} \quad . \quad (10.30)$$

To arrive at the admittance matrix for the actual oscillator, we simply superimpose upon $\mathbf{Y}'_n$ (*see* Fig. 10.6(*c*) the matrix of the triode (eqn. (10.25)). This gives us

$$\mathbf{Y}' = \mathbf{Y}'_n + \mathbf{Y}'_v = \begin{Bmatrix} (R + j\omega L_2)/D & j\omega M/D \\ j\omega M/D + g_m & j\omega L_1/D + j\omega C + k \end{Bmatrix} \quad . \quad (10.31)$$

If oscillations are to develop, det ($\mathbf{Y}'$) must vanish. Hence the complex equation

$$\det(\mathbf{Y}') = \{1 + (k + j\omega C)(R + j\omega L_2) - j\omega M g_m\}/D = 0 \quad . \quad (10.32)$$

must be satisfied. D is always non-zero and can therefore be ignored. Equating real and reactive members, we find

$$M = (RC + kL_2)/g_m \quad . \quad . \quad . \quad . \quad (10.33)$$

which shows how the coupling necessary to bring about the oscillatory condition depends on the circuit parameters and the valve constants, and

$$1 + kR - \omega^2 L_2 C = 0 \quad . \quad . \quad . \quad . \quad (10.34)$$

which implies that oscillations can occur only at the frequency

$$f = \frac{1}{2\pi}\sqrt{\frac{1 + kR}{L_2C}} \quad . \quad . \quad . \quad . \quad (10.35)$$

In practice the amplitude of oscillation rapidly increases until its further growth is limited by non-linear effects which were not taken into account in the above analysis.

10.5. Algebraic Diagram of Network Analysis

In the field of algebraic topology, extensive use is made of diagrams to show the interrelation of algebraic entities. This technique can be successfully applied to demonstrate the algebraic pattern of network analysis as developed in Sections 10.3 and 10.4. The resulting diagram is given in Fig. 10.7.

From this diagram all the transformations derived in this chapter can be read. It should be borne in mind that e and I are b-dimensional, active, independent vector variables, whereas E and i are

Node-pair Branch Mesh

$$\begin{tikzcd}
E' \arrow[r, "A"] \arrow[d, shift right, "A_tYA"'] & V=E+e \arrow[r, "C_t"] \arrow[d, shift right, "Y"'] & e' \arrow[d, shift left, "(C_tZC)^{-1}"] \\
I' \arrow[u, shift right, "(A_tYA)^{-1}"'] & J=I+i \arrow[l, "A_t"'] \arrow[u, shift right, "Z"'] & i' \arrow[l, "C"'] \arrow[u, shift left, "C_tZC"]
\end{tikzcd}$$

FIG. 10.7. ALGEBRAIC DIAGRAM OF NETWORK ANALYSIS

response quantities that are constrained by having to obey Kirchhoff I and II respectively.

An arrow in the diagram denotes pre-multiplication by the associated symbol. Thus the arrow alongside $\mathbf{A}_t$ takes J into I':

$$I' = \mathbf{A}_tJ = \mathbf{A}_tI + \mathbf{A}_ti = \mathbf{A}_tI + 0 = \mathbf{A}_tI \qquad . \quad (10.11)$$

Further pre-multiplication by $(\mathbf{A}_t\mathbf{Y}\mathbf{A})^{-1}$ and, subsequently, by $\mathbf{A}$, gives us

$$E = \mathbf{A}(\mathbf{A}_t\mathbf{Y}\mathbf{A})^{-1}\mathbf{A}_tI \quad . \quad . \quad . \quad (10.22)$$

It is worth noting that $\mathbf{C}$, $\mathbf{C}_t$, $\mathbf{A}$ and $\mathbf{A}_t$ are associated with arrows in one direction only; this is because they represent singular matrices ($r(\mathbf{C}) = m$ and $r(\mathbf{A}) = p$) that cannot be inverted.

Adopting a geometric point of view, it is apparent that the middle part of the diagram, which refers to branch quantities, requires a b-dimensional space for its presentation.

The node-pair end of the diagram is p-dimensional and voltage vectors in this space are mapped into b-space by means of the non-invertible matrix $\mathbf{A}$. They are confined to the p-dimensional subspace spanned by the column vectors of $\mathbf{A}$, and this subspace is orthogonal to the m-dimensional "mesh-space" spanned by the column vectors of $\mathbf{C}$ as evidenced by eqn. (10.13).

At the right-hand end of the diagram we have the m-dimensional mesh quantities e' and i'. $\mathbf{C}$ maps the independent mesh-currents i' into an m-dimensional subspace of b-space.

The node-pair and mesh spaces are complementary (together they fill all b-space), and this fact is the basis of an approach to network analysis which deals simultaneously with node-pair and mesh quantities. We shall discuss the *orthogonal* or *mesh-node method* in the next section.

Exercise 1. If we limit the excitation of a network to series-injected e.m.f.s e only, and let e vary throughout all b-space, will the response immittance voltages V (and hence also the immittance currents J) cover all b-space?

From eqn. (10.14), when $I = 0$, we have

$$V = E + e = \mathbf{Z}i \quad . \quad . \quad . \quad . \quad (10.36)$$

Superficially, it might look as though V would sweep through all b-space if e were allowed to vary freely. It should not be forgotten, however, that E is a function of e as well as of I.

Substituting eqn. (10.18) in eqn. (10.36), we see that V can be expressed in terms of the network constants and e as

$$V = \mathbf{Z}\mathbf{C}(\mathbf{C}_t\mathbf{Z}\mathbf{C})^{-1}\mathbf{C}_te \quad . \quad . \quad . \quad . \quad (10.37)$$

from which it is apparent that the V will only cover an m-dimensional subspace of b-space.

Speaking very loosely, eqn. (10.37) could be described as follows: The operator $\mathbf{C}_t$ squeezes the free b-dimensional vector e into m dimensions; $(\mathbf{C}_t\mathbf{Z}\mathbf{C})^{-1}$ transforms it into its counterpart in the current space without changing its dimensionality; $\mathbf{C}$ injects this current space into b-space; and, finally, the square matrix $\mathbf{Z}$ carries it back into the original b-dimensional voltage space. By consulting Fig. 10.7, it can be seen that we have moved in a clockwise direction round the right-hand loop of the algebraic diagram.

Analogously, if we make $e = 0$ and express V as a function of I only, we arrive at the expressions

$$V = \mathbf{A}(\mathbf{A}_t\mathbf{Y}\mathbf{A})^{-1}\mathbf{A}_tI \quad . \quad . \quad . \quad . \quad (10.38)$$

and

$$J = \mathbf{Y}\mathbf{A}(\mathbf{A}_t\mathbf{Y}\mathbf{A})^{-1}\mathbf{A}_tI \quad . \quad . \quad . \quad . \quad (10.39)$$

where V and J will cover a p-space when I varies freely.

On account of the linearity of the network, the simultaneous application of e and I will result in

$$V = \mathbf{Z}\mathbf{C}(\mathbf{C}_t\mathbf{Z}\mathbf{C})^{-1}\mathbf{C}_te + \mathbf{A}(\mathbf{A}_t\mathbf{Y}\mathbf{A})^{-1}\mathbf{A}_tI. \quad . \quad . \quad (10.40)$$

and

$$J = \mathbf{C}(\mathbf{C}_t\mathbf{Z}\mathbf{C})^{-1}\mathbf{C}_te + \mathbf{Y}\mathbf{A}(\mathbf{A}_t\mathbf{Y}\mathbf{A})^{-1}\mathbf{A}_tI. \quad . \quad . \quad (10.41)$$

As a further exercise the reader should test how these two equations react to pre-multiplication by $\mathbf{C}_t$ and $\mathbf{A}_t$.

Exercise 2. If we write eqn. (10.40) in the form

$$V = \mathbf{Z}\mathbf{C}(\mathbf{C}_t\mathbf{Z}\mathbf{C})^{-1}e' + \mathbf{A}(\mathbf{A}_t\mathbf{Y}\mathbf{A})^{-1}I' \quad . \quad . \quad . \quad (10.42)$$

it becomes clear that the voltage drops across all the branch immittances and the currents flowing through them will be invariant, provided that the independent mesh e.m.f.s e', and the independent node-pair currents I', remain unaltered.

10.6. Orthogonal Analysis (Mesh-Node Analysis)

Kron was the first to show the way to a system of analysis combining the mesh and the node-pair methods. He probably had eqn. (10.13) in mind when he called his method "orthogonal analysis." In our discussion we shall sometimes use the words *mesh-node* analysis to mean the same thing. Furthermore, we shall construct a canonical (basic) set of mesh-node connexion matrices (**A** and **C**) by selecting the independent node-pair p.d.s E' and the independent mesh currents i' as follows.

To begin with, we select any set of independent node-pairs such that each member of the set is the p.d. E'_i across a branch. This set must define a tree on the network, for if it did not, we should be able to pick out a closed loop round which we should have

$$\Sigma E'_i = 0$$

thus indicating that the E'_i's were dependent, contrary to our hypothesis.

As the independent meshes we choose the basic meshes corresponding to the cotree of the tree defined by the E'_i.

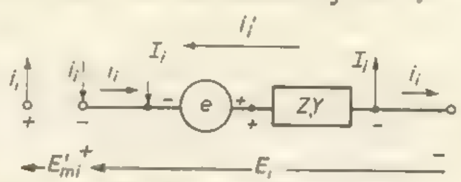


Fig. 10.8. Opening a Basic Mesh

The dual of this method of selection is also possible as discussed in Exercise 1 below.

Consider now the node-pair connexion matrix **A** with one row for each branch of the network and one column for each independent node-pair (each branch of the tree on the network). If we were able to open the m branches of the cotree without interrupting the mesh currents flowing through them, we should create m new node-pairs and thus be able to augment **A** by adding m columns, thus making the matrix square. This can be done by injecting m currents as illustrated in Fig. 10.8.

The p.d. E'_{mi} across the break in the mesh is zero, and the operation will not in any way affect the currents and voltages of the network.

Thus, the b independent node-pair p.d.s will consist of p basic node-pair voltages E'_{pi} across the branches of the tree plus m node-pair voltages E'_{mi} (= 0) in series with the branches of the cotree.

The branch p.d.s can be expressed as linear combinations of E'_p and E'_m by means of the matrix equation

$$E = \begin{Bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{L} & -\mathbf{I} \end{Bmatrix} \begin{Bmatrix} E'_p \\ E'_m \end{Bmatrix} = (\mathbf{A}_p \mathbf{A}_m) \begin{Bmatrix} E'_p \\ E'_m \end{Bmatrix} = \mathbf{A} \begin{Bmatrix} E'_p \\ E'_m \end{Bmatrix} = \mathbf{A}E' \quad (10.43)$$

where the b-by-p compound matrix $\begin{Bmatrix} \mathbf{I} \\ \mathbf{L} \end{Bmatrix} = \mathbf{A}_p$ is identical with the node-pair connexion matrix **A** of eqn. (10.10).

It is readily verified that **A** is its own inverse—

$$\mathbf{A}^2 = \mathbf{I} \quad . \quad . \quad . \quad . \quad (10.44)$$

and also that the determinant of **A** is

$$\det(\mathbf{A}) = (-1)^m \quad . \quad . \quad . \quad (10.45)$$

Before compiling the mesh-node matrix corresponding to the mesh connexion matrix **C** of eqn. (10.6), let us study the m-by-p submatrix **L** more closely.

Each row of **L** corresponds to a branch of the cotree, and each column to a branch of the tree on the network.

The elements in row k of **L** (which are +1, −1 or 0) indicate and orient the branches of the tree that combine with link k of the cotree to form the kth basic mesh. At the same time, k shows us which basic node-pair p.d.s E'_{pi} are used, together with E'_{mk}, to calculate E_k.

From this it follows that the elements in a column of **L** indicate (with orientation) the basic mesh currents passing through the corresponding branch of the tree.

In order to augment the mesh connexion matrix so as to make it square, we adopt a technique which is the dual of that employed to augment **A**. Instead of opening a branch of the cotree by injecting the branch current i from an admittanceless current generator, we short-circuit a branch of the tree by applying a p.d. E from an impedanceless source of e.m.f. across the branch. No current will flow in the artificial mesh created because the existing p.d. across the branch and the applied voltage are equal and opposed, thus cancelling each other. The orientation of the mesh current i'_{pi} (= 0) coincides with that of i_i, as shown in Fig. 10.9.

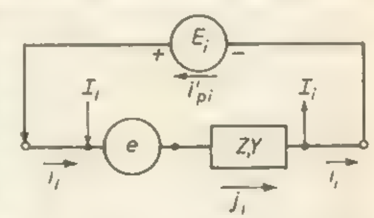


Fig. 10.9. Short-circuiting a Basic Node-pair

Referring to our remarks about the matrix $\mathbf{L}$, we find that the branch currents i are given as linear combinations of the basic mesh currents i'_m and the artificial mesh currents i'_p by means of the equation

$$i = \begin{Bmatrix} \mathbf{I} & \mathbf{L}_t \\ \mathbf{O} & -\mathbf{I} \end{Bmatrix} \begin{Bmatrix} i'_p \\ i'_m \end{Bmatrix} = (\mathbf{C}_p \mathbf{C}_m) \begin{Bmatrix} i'_p \\ i'_m \end{Bmatrix} = \mathbf{C} \begin{Bmatrix} i'_p \\ i'_m \end{Bmatrix} = \mathbf{C}i' \qquad (10.46)$$

where $\mathbf{C}_m = \begin{Bmatrix} \mathbf{L}_t \\ -\mathbf{I} \end{Bmatrix}$ is the mesh connexion matrix denoted by $\mathbf{C}$ in Section 10.2.

$\mathbf{C}$ is its own inverse and is equal to the transpose of $\mathbf{A}$. Thus

$$\mathbf{C}^2 = \mathbf{C}\mathbf{A}_t = \mathbf{A}_t\mathbf{C} = \mathbf{I} \qquad (10.47)$$

and hence

$$\det(\mathbf{C}) = \det(\mathbf{A}) = (-1)^m \qquad (10.48)$$

Substituting the compound forms of $\mathbf{A}$ and $\mathbf{C}$ from eqns. (10.43) and (10.46) in eqn. (10.47), we get

$$\begin{aligned}(\mathbf{C}_p\mathbf{C}_m)_t(\mathbf{A}_p\mathbf{A}_m) &= \begin{Bmatrix} \mathbf{C}_{pt} \\ \mathbf{C}_{mt} \end{Bmatrix} (\mathbf{A}_p\mathbf{A}_m) \\ &= \begin{Bmatrix} \mathbf{C}_{pt}\mathbf{A}_p & \mathbf{C}_{pt}\mathbf{A}_m \\ \mathbf{C}_{mt}\mathbf{A}_p & \mathbf{C}_{mt}\mathbf{A}_m \end{Bmatrix} = \begin{Bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{Bmatrix} \qquad (10.49)\end{aligned}$$

The identity $\mathbf{C}_{mt}\mathbf{A}_p = \mathbf{0}$, which is implicit in this equation is the same as eqn. (10.13).

Exercise 1. In this exercise we shall prove that it is always possible to choose an independent set of meshes in a network in such a way that each mesh current passes through at least one branch which is not linked by any other mesh. This set of branches forms a cotree on the network.

Let us consider a branch b_1 connecting nodes n_1 and n_2. If b_1 is linked by more than one mesh, we shall reroute all but one (say m_1) of the meshes through other paths connecting n_1 and n_2. This is equivalent, algebraically, to cancelling all the non-zero elements in row b_1 of the mesh-connexion matrix $\mathbf{C}$ by means of linear operations with column m_1. Such a process does not alter the rank of $\mathbf{C}$, and hence the modified meshes will still be independent.

Repeating this procedure, we can pair off the m independent meshes of the network with m of its branches; the remaining $b - m = p$ branches must form a tree on the network; for, if they did not, they would enclose a mesh which could be associated with one of its branches.

Exercise 2. Applying the mesh-node technique to the circuit of Fig. 10.3 (Exercise 1, Section 10.2), we choose branches 1, 3, 4 as a tree on the network; branches 2, 5, 6, then, are the links. In order to conform to the canonical method, branches 1, 3, 4, 2, 5, 6 are renumbered 1′, 2′, 3′, 4′, 5′, 6′.

This results in an orthogonal mesh matrix—

$$\mathbf{C} = \begin{Bmatrix} 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{Bmatrix}$$

and the corresponding orthogonal node-pair matrix

$$\mathbf{A} = \begin{Bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 & 0 & -1 \end{Bmatrix} = \mathbf{C}_t$$

The reader should verify by direct multiplication that $\mathbf{C}_t\mathbf{A} = \mathbf{I}$.

Branch and mesh e.m.f.s are linked by the matrix equation

$$e' = \mathbf{C}_t e \qquad (10.50)$$

If we write this expression out in full, it can be seen that the artificial e.m.f.s applied across the branches of the tree do not appear in the expanded formula. They are in reality unknown response quantities, and not, like the real branch e.m.f.s e_l, independent variables.

An identity analogous to eqn. (10.11) is also valid—

$$I' = \mathbf{A}_t I \qquad (10.51)$$

All these relations between the network variables and the connexion matrices are summarized in Table 10.2.

Table 10.2

ORTHOGONAL RELATIONS BETWEEN NETWORK VARIABLES

$E = \mathbf{A}E'$	$E' = \mathbf{C}_t E$
$i = \mathbf{C}i'$	$i' = \mathbf{A}_t i$
$I' = \mathbf{A}_t I$	$I = \mathbf{C}I'$
$e' = \mathbf{C}_t e$	$e = \mathbf{A}e'$

$$\mathbf{A}_t\mathbf{C} = \mathbf{C}\mathbf{A}_t = \mathbf{I}$$

Exercise 3. To make sure that the orthogonal transformations of the network variables leave the total dissipated power invariant, we compute the instantaneous power dissipated in the branch immittances—

$$\begin{aligned} P = V_t J = (E_t + e_t)(I + i) &= E_t I + E_t i + e_t I + e_t i \\ &= E'_t \mathbf{A}_t \mathbf{C} I' + E'_t \mathbf{A}_t \mathbf{C} i' + e'_t \mathbf{A}_t \mathbf{C} I' + e'_t \mathbf{A}_t \mathbf{C} i' \\ &= (E'_t + e'_t)(I' + i') = P' \qquad (10.52)\end{aligned}$$

To derive the network equations for a network in the orthogonal manner, we start as before (*see* eqn. (10.14)) with Ohm's law—

$$E + e = \mathbf{Z}(I + i) \qquad (10.53)$$

Pre-multiplication by $\mathbf{C}_t$ and the insertion of the factor $\mathbf{CC} = \mathbf{I}$ between $\mathbf{Z}$ and $\boldsymbol{I} + \boldsymbol{i}$ gives us (with the help of the identities in Table 10.2)

$$\boldsymbol{E}' + \boldsymbol{e}' = \mathbf{C}_t\mathbf{ZC}(\boldsymbol{I}' + \boldsymbol{i}') \quad . \quad . \quad . \quad . \quad (10.54)$$

and

$$\boldsymbol{I}' + \boldsymbol{i}' = \mathbf{A}_t\mathbf{YA}(\boldsymbol{E}' + \boldsymbol{e}') \quad . \quad . \quad . \quad . \quad (10.55)$$

Here $(\mathbf{A}_t\mathbf{YA})^{-1} = \mathbf{C}_t\mathbf{ZC}$, since

$$(\mathbf{A}_t\mathbf{YA})(\mathbf{C}_t\mathbf{ZC}) = \mathbf{A}_t\mathbf{Y}(\mathbf{AC}_t)\mathbf{ZC} = \mathbf{A}_t(\mathbf{YZ})\mathbf{C} = \mathbf{A}_t\mathbf{C} = \mathbf{I} \qquad \text{(Q.E.D.)}$$

To be more explicit, we can write $\mathbf{C}_t\mathbf{ZC}$ and $\mathbf{A}_t\mathbf{YA}$ as follows—

$$\mathbf{C}_t\mathbf{ZC} = \begin{Bmatrix} \mathbf{C}_{pt}\mathbf{ZC}_p & \mathbf{C}_{pt}\mathbf{ZC}_m \\ \mathbf{C}_{mt}\mathbf{ZC}_p & \mathbf{C}_{mt}\mathbf{ZC}_m \end{Bmatrix} \quad . \quad . \quad . \quad (10.56)$$

and

$$\mathbf{A}_t\mathbf{YA} = \begin{Bmatrix} \mathbf{A}_{pt}\mathbf{YA}_p & \mathbf{A}_{pt}\mathbf{YA}_m \\ \mathbf{A}_{mt}\mathbf{YA}_p & \mathbf{A}_{mt}\mathbf{YA}_m \end{Bmatrix} \quad . \quad . \quad . \quad (10.57)$$

Comparison of eqn. (10.56) with eqn. (10.15), and of eqn. (10.57) with eqn. (10.22), will convince us that the bottom right-hand matrix element of eqn. (10.56) is the mesh impedance matrix, and the top left-hand matrix element of eqn. (10.57), the node-pair admittance matrix of the network.

Exercise 4. The mesh-node connexion matrices developed and discussed in this section were shown to have determinants equal to $(-1)^m$. It is not necessary, however, to employ basic node-pairs and meshes in order to obtain square non-singular connexion matrices. If we compute a mesh-node connexion matrix in any other way, its columns (rows) will be independent linear combinations of the columns (rows) of the basic matrix. Hence, we can state as a general theorem that the determinant of a mesh-node connexion matrix must be equal to either $+1$ or -1.

To conclude this chapter we shall define and briefly discuss the *duality* of *planar networks.*

A planar network is one whose graph can be drawn on a plane (or on the surface of a sphere) without any of the branches crossing each other. From a purely geometric point of view a planar graph can be considered as a complex of nodes, branches and faces. Each face is bounded by a set of branches forming a mesh, and each branch is bound by a set of (two) nodes forming a node-pair. These entities are conveniently ordered in a sequence—

Node—Node-pair—Branch—Mesh—Face

which is symmetric about its middle term.

We can exploit this symmetry to establish a relationship, termed *topological duality*, between planar networks such that—

(i) Each node in one network corresponds to a face in the other, and vice versa.

(ii) Each branch in one network corresponds to a branch in the other.

(iii) The branches connected to a node in one network correspond to the branches bounding a face (i.e. forming a mesh) in the other, and vice versa.

The diagram in Fig. 10.7 can be enlarged by adding a dot representing "node" to the left of the dot marked "node-pair," and a dot marked "face" at the extreme right of the diagram. This modified diagram suggests that the duality relationship is represented by the symmetry of the algebraic diagram. It also gives a hint as to further properties of networks. First, that a non-planar network in which certain of the meshes do not bound a face, and where the diagram consequently loses its symmetry, does not possess topological duals. Secondly, that because neither a face nor a single node has any electrical significance, the diagram of Fig. 10.7 will be symmetric even for non-planar networks, and a network therefore always has an electrical, if not a topoplogical, dual.

We shall now proceed to formulate and prove a relationship between the trees and cotrees of dual networks.

Theorem 42. The dual of a tree on a planar network is a cotree on the dual network, and vice versa.

Proof. Let N_1 and N_2 be two planar, connected, dually related networks. A tree T_1 on N_1 is, by definition, a graph connecting all the n_1 nodes of N_1 without forming any meshes. T_1, therefore, does not divide the plane into disjoint regions, and any two points of the plane can be connected without having to cross a branch of T_1.

The meshes of N_1 are now so chosen that each mesh encloses a face. A mesh of this type is called a *window mesh* or merely a *window*. The enveloping mesh, which is a linear combination of the independent meshes of the network, can be said to enclose the region lying outside it. This terminology becomes far less paradoxical if we think of a planar network drawn on the surface of a sphere. If N_1 has m_1 independent meshes, we can plot $n_2 = m_1 + 1$ nodes in the plane, each node corresponding to a mesh of N_1 (including the enveloping mesh).

Each of the nodes of N_2 is thus enclosed by a mesh of N_1 which contains at least one branch of the cotree T_1' on N_1. Also, the $n_2 = m_1 + 1$ nodes of N_2 can be connected by branches which do not intersect T_1, and the graph formed in this way cannot contain any closed loops; for, if it did, the loop would enclose a node of N_1, which would not, therefore, form part of T_1, thus contradicting our hypothesis that T_1 is a tree on N_1.

The graph comprising $p_2 = n_2 - 1 = m_1$ branches connecting all n_2 nodes of N_2, without forming any loop, is therefore a tree T_2 on N_2.

To complete the proof, we can pursue an analogous train of reasoning to show that the dual of the cotree T_2^* on N_2 is a tree T_1 on N_1.

In view of this theorem, it is to be expected that there must exist a close relation between the basic connexion matrices of two dually related planar networks. This relationship is given in the next theorem. Before stating the theorem, however, we must decide on a system of defining the relative orientation of the branches of two planar, dual networks.

We shall adopt the following procedure: After orienting the branches of (say) network N_1 arbitrarily, we orient a branch b_2 forming part of a window mesh m_2 of N_2 in a clockwise or an anti-clockwise direction round m_2, according to whether the image b_1 of b_2 is directed towards or away from the node n_1 of N_1 which is the image of mesh m_2.

An interchange of the terms *clockwise* and *anti-clockwise* results in a reversal of orientation of all the branches of N_2, an operation which leaves the connexion matrices invariant.

Once the branches of the two networks have been given an orientation, the orientations of the injected currents and voltages and of the basic mesh currents and node-pair voltages are determined by the polarity conventions indicated in Figs. 10.8 and 10.9.

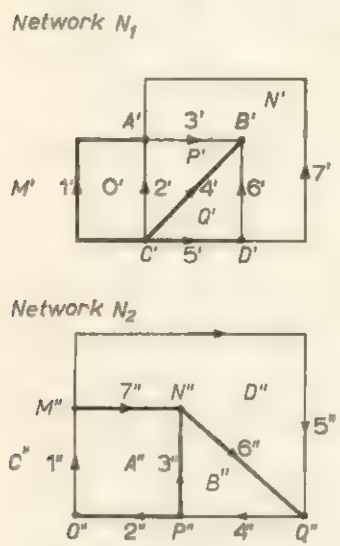


FIG. 10.10. GRAPHS OF DUAL PLANAR NETWORKS

As an example of the method described above, take node D' of network N_1, Fig. 10.10. Of the three branches connected to D', branch $5'$ converges, whereas branches $6'$ and $7'$ diverge. Hence, in N_2, where branches $5''$, $6''$ and $7''$ complete mesh D'', branch $5''$ is oriented in a clockwise, branches and $6''$ and $7''$ are oriented in an anti-clockwise, direction round the mesh.

Theorem 43. The basic node-pair connexion matrix of a network is equal to the negative of the cotranspose of the basic mesh connexion matrix of its dual, where the operation *cotransposition* denotes transposition about the second diagonal.

The proof follows readily from Theorem 42 when we recall that the duality relation maps branches converging on a node into branches forming a mesh, a tree into a cotree, and vice versa, and compare this with the significance of the rows and columns of the basic connexion matrices (*see* the discussion of submatrix **L** of the basic node-pair connexion matrix **A**, at the beginning of this section).

The operation of cotransposition, which is equivalent to renumbering the branches in reverse order, is necessary so as to place the branches of the tree of the dual network in the correct position in the basic connexion matrix, as required by our canonical method of analysis.

Exercise 5. Fig. 10.10 illustrates the concept of duality between two planar networks N_1 and N_2. To emphasize the symmetry of the relationship, the characteristic constants (number of branches, number of independent meshes, etc.) of the two networks are given below in tabular form—

$$\begin{array}{rcl} N_1 & & N_2 \\ \hline b_1 = & 7 & = b_2 \\ s_1 = & 1 & = s_2 \\ p_1 = n_1 - s_1 = & 3 & = m_2 \\ m_1 = & 4 & = n_2 - s_2 = p_2 \\ \hline \end{array}$$

Note how the branches converging on a node in one network (say $3'$, $4'$, $6'$ of N_1) correspond to branches forming a mesh ($3''$, $4''$, $6''$ of N_2) in the other network, and vice versa.

We can thus superimpose N_1 on N_2 in such a way that node A' will lie in mesh A'', node B' in mesh B'', and so forth. Mesh Q' will then enclose node Q'', etc.

The branches $1'$, $4'$ and $5'$ form a tree on N_1. Hence, by Theorem 42, the branches $2''$, $3''$, $6''$ and $7''$ of N_2 must form a tree on N_2. This is readily verified.

Exercise 6. If we choose branches $1'$, $4'$ and $5'$ as a tree on N_1 of Exercise 5, the mesh-node connexion matrix $\mathbf{A}'$ of N_1 becomes

$$\mathbf{A}' = \begin{array}{c} \\ 1' \\ 4' \\ 5' \\ 2' \\ 3' \\ 6' \\ 7' \end{array} \begin{array}{c} \begin{array}{ccccccc} 1' & 4' & 5' & 2' & 3' & 6' & 7' \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & -1 \end{bmatrix} \end{array} = \mathbf{C}'_t$$

Notice how the rows of $\mathbf{A}'$ indicate (with sign) the branch p.d.s of the covering tree that combine to give the p.d. across the branch in question. Branch $4'$, for example, being part of the tree on N_1, requires only one coordinate E'_{p4}. The p.d. across branch $6'$, on the other hand, is a linear combination of the voltages E'_{p4}, E'_{p5} and E'_{m6} $(= 0)$.

Turning our attention to the mesh axes of the network, it is readily seen by inspection that the rows of $\mathbf{C}'$ (which are the same as the columns of $\mathbf{A}'$, since $\mathbf{C}' = \mathbf{A}'_t$) indicate (with orientation) the meshes linking each branch. Each branch of the cotree (say 2′) carries only one basic mesh current (i'_{m2}), whereas a branch of the tree (say 5′) may carry one or more basic mesh currents (i'_{m6} and i'_{m7}) in addition to its own artificial mesh current ($i'_{p5} = 0$).

According to Theorem 43 the basic node-pair connexion matrix of N_2 is equal to the negative of the cotranspose of $\mathbf{A}'$.

Hence,

$$\mathbf{A}'' = -\mathbf{A}'_k = \begin{array}{c} \\ 7'' \\ 6'' \\ 3'' \\ 2'' \\ 5'' \\ 4'' \\ 1'' \end{array} \!\!\begin{array}{c} \begin{array}{ccccccc} 7'' & 6'' & 3'' & 2'' & 5'' & 4'' & 1'' \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & -1 & 0 \\ -1 & 0 & 1 & -1 & 0 & 0 & -1 \end{bmatrix} \end{array}$$

as can readily be verified.

Exercise 7. If cotransposition is indicated by a subscript k, show that

$$(\mathbf{AB})_k = \mathbf{B}_k\mathbf{A}_k \quad \text{. . . . . (i)}$$
$$(\mathbf{A}_t)_k = (\mathbf{A}_k)_t (= \mathbf{A}_{tk} = \mathbf{A}_{kt}) \quad \text{. . . . . (ii)}$$
$$(\mathbf{A}^{-1})_k = (\mathbf{A}_k)^{-1} = \mathbf{A}_k^{-1} \quad \text{. . . . . . (iii)}$$
$$\det(\mathbf{A}_k) = \det(\mathbf{A}) \quad \text{. . . . . . . (iv)}$$

In other words, prove that cotransposition follows the same rules as transposition and that the two operations commute; also, show that cotransposition and inversion commute, and that the determinant of a matrix is invariant to cotransposition.

PROBLEMS

10.1. Let a network with four nodes, A, B, C and D, be given. The network comprises five branches, AB, BC, AC, CD and AD. Employ Euler's theorem (Theorem 41) to determine the number of independent meshes in the network.

Are the meshes ABC, ACD and $ABCD$ independent? Orient these meshes, and express the branch currents as linear combinations of the mesh currents by means of a 5-by-3 connexion matrix $\mathbf{C}$. Determine the rank of $\mathbf{C}$.

10.2. Draw a tree on each of the following networks: (i) a Wheatstone bridge, (ii) a network consisting of five branches in parallel, (iii) a network consisting of five branches connected in series to form one mesh, (iv) a network consisting of six branches connected in star.

10.3. Show that, if eqn. (10.7) is written $P = \boldsymbol{e}_t\boldsymbol{i} = \boldsymbol{e}'_t\boldsymbol{i}'$, we arrive at the same equation as before for $\boldsymbol{e}'$ in terms of $\boldsymbol{e}$.

10.4. Draw graphs of all the networks that can be constructed from three branches, and write down in each case the corresponding orthogonal connexion matrices.

10.5. Prove by direct matrix multiplication that expressions (10.56) and (10.57) are the inverses of each other. *Hint.* Employ the identities $\mathbf{CA}_t = \mathbf{AC}_t = \mathbf{C}_t\mathbf{A} = \mathbf{A}_t\mathbf{C} = \mathbf{I}$, where $\mathbf{A} = (\mathbf{A}_p\mathbf{A}_m)$ and $\mathbf{C} = (\mathbf{C}_p\mathbf{C}_m)$.

10.6. Draw the graph of any planar network (N_1) and construct the graph of its dual (N_2). Choose a tree T_1 on N_1, and show that the dual of T_1 is a cotree, T'_2 on N_2.

10.7. Compile the basic node-pair matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ of N_1 and N_2 respectively (*see* Problem 10.6), and test the validity of Theorem 43, which requires that $\mathbf{A}_2 = -(\mathbf{A}_1)_k$.

10.8. Calculate the node-pair connexion matrix of a network which is a tree on itself (an all-node-pair network). What is the dual of such a network, and what is its basic node-pair connexion matrix?

10.9 Prove that $\mathbf{A}_k = (\mathbf{UAU})_t = \mathbf{UA}_t\mathbf{U}$, where $\mathbf{U} = (U_n \ldots U_2U_1)$.

10.10. Prove that $\det(\mathbf{U}) = (-1)^{\frac{1}{2}n(n-1)}$, where $\mathbf{U}$ is the square matrix defined in Problem 10.9. Prove that $\mathbf{U}^2 = \mathbf{I}$. Use these results to prove the statements made in Exercise 7, Section 10.6.

10.11. Compile the mesh connexion matrix and the node-pair connexion matrix of the network shown in Fig. 10.11.

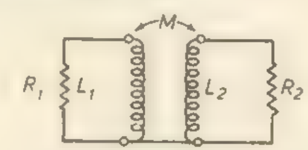


FIG. 10.11

10.12. Supply a detailed proof of eqn. (10.9) by expanding the branch currents in terms of their mesh current components and then factorizing the mesh currents.

10.13. Prove eqn. (10.11), $\boldsymbol{I}' = \mathbf{A}_t\boldsymbol{I}$, by requiring that the total power $P = \boldsymbol{I}_t\boldsymbol{E}$ fed to a network be invariant to a change from a branch to a node-pair point of view.

10.14. Verify eqns. (10.43) and (10.46) in detail. Note in particular how the choice of orientation of E'_{mt} and i'_{pt} leads to the submatrix $-\mathbf{I}$ in the connexion matrices $\mathbf{A}$ and $\mathbf{C}$, and how this, in turn, ensures that the simple canonical eqns. (10.44) and (10.47) are valid.

10.15. Discuss how, in a planar, connected network, the enveloping mesh is dually equivalent to the redundant node (which can be earthed).

10.16. Draw a number of oriented graphs, compile the connexion matrices in each case, and verify by inspection the functions of $\mathbf{C}_t$ and $\mathbf{A}_t$ as pre-factors (*see* Table 10.1).

10.17. Let a bridge network such as the one in Fig. 10.3 be given and let the conductances of the branches be as follows—

Branch	1	2	3	4	5	6
Conductance (mhos) . . .	1	2	1	3	2	4

Calculate the node-pair admittance matrix $\mathbf{Y}'$ assuming that no couplings exist between the branches.

Suppose now that the following currents are injected at the nodes of the network—

Node	0	1	2	3
Current, I_i (amperes)	−1	2	3	−4

Is this system of injected currents physically realizable?

Express the injected currents in terms of three independent sets of currents $+I_i$ and $-I_i$ ($i = 1, 2, 3$) across the independent node-pairs that were used to compile the connexion matrix A (*see* Exercise 1, Section 10.2).

Calculate the p.d.s E_i appearing across the branches of the network.

CHAPTER 11

Diakoptics

11.1. Introduction

Of all the matrix operations defined in this book, that of inverting a square non-singular matrix is by far the most laborious. Any new technique, such as diakoptics, designed to facilitate this process is therefore worthy of careful consideration.

In the preceding chapter we have shown that the solution of a linear network problem requires *inter alia* the inversion of an m-by-m matrix when the network is analysed as a mesh network, and a p-by-p matrix when the node-pair method is employed.

The amount of labour required to invert a matrix increases roughly with the third power of its dimensions so that, even when an automatic computer is available, the task rapidly becomes prohibitively time-consuming, quite apart from the fact that there is a limit to the size of the matrices which any given computer can invert.

Kron's method of *diakoptics*, or *tearing*, greatly reduces the dimensions of the matrices to be inverted by dividing the system into a number of smaller subsystems. Each subsystem is then solved as an independent entity, and the solutions obtained are combined to form the required solution of the total system. In order to interconnect the partial solutions it is necessary to invert an additional *tie matrix* (*intersection matrix*), but in spite of this, the net result is a considerable reduction in the total amount of computational labour, not to mention that the diakoptic technique can bring a problem within the range of an available computer which might otherwise have been incapable of handling the problem at all.

A rough calculation will illustrate the saving in time. Let the time required to invert an N-by-N matrix be $T = kN^3$, and let us assume that the network can be divided into n subsystems of approximately equal size. Then each subsystem will be represented by an N/n-by-N/n matrix and the total time required to solve the system diakoptically will be of the order of

$$T_d = 2kn(N/n)^3 = T/(\tfrac{1}{2}n^2)$$

where a safety factor of 2 has been included to taken into account the additional labour required for the division into subsystems and

the inversion of the tie impedance matrix, and for the computation of various additional products. As can be seen from this formula, even a division into two or three parts will yield an appreciable reduction in labour.

According to Kron's system of *tearing*, the division of a given network is performed by detaching suitable tie branches so as to create a number of unconnected subnetworks. To compensate for the branches removed by this process, currents equal to those that were flowing through the branches are injected at the nodes from which they were disconnected. This method, based on the node-pair approach, is called *diakoptics*.

The dual method achieves the same division by applying short-circuits. This method, described by Onodera, has been given the name *codiakoptics*. We shall, however, confine ourselves to diakoptic analysis.

As a necessary preliminary, we shall study the mesh impedance matrix and the node-pair admittance matrix of a network in greater detail.

Assuming a given network to be energized solely by series sources of e.m.f., e, the mesh equation of the circuit is

$$e' = \mathbf{C}_t\mathbf{Z}\mathbf{C}i' = \mathbf{Z}'i' \qquad . \quad . \quad . \ (11.1)$$

In addition, we make the assumption that no couplings exist between the branches of the network. The primitive impedance matrix of the network is therefore diagonal with the general element

$$z_{ij} \begin{cases} = 0 & \text{when } i \neq j \\ \neq 0 & \text{when } i = j \end{cases}$$

Also, the general element of C is

$$c_j^i \begin{cases} = 0 & \text{when branch } i \text{ and mesh } j \text{ are not incident} \\ = \pm 1 & \text{when branch } i \text{ and mesh } j \text{ are positively/} \\ & \quad \text{negatively incident} \end{cases}$$

The mesh impedance matrix $\mathbf{Z}'$ can be written

$$c_u^i z_{uv} c_j^v = z'_{ij} \qquad (u, v = 1, \ldots, b) \qquad . \quad . \ (11.2)$$

from which we deduce that z'_{ij} is equal to the sum of the impedances common to meshes i and j, each impedance being multiplied by ± 1, depending on whether mesh i and mesh j are positively or negatively incident. Note that the orientation of an individual branch has no effect on the sign of the elements of the relevant rows and columns of $\mathbf{Z}'$.

A mesh can be considered to be positively incident to itself, and thus element z'_{kk} is the sum of the impedances round mesh k.

Because this method of writing down the mesh impedance matrix for a network without inter-branch couplings is so familiar, we have handled its proof rather superficially. The dual method of compiling the node-pair matrix of a network by inspection is not universally known, and we shall proceed with more circumspection.

To make our derivation more convincing, we shall base our arguments directly on the network without invoking the matrix formulae of Section 10.4. Consider, therefore, a node n_1 of the network, and let the injected current at n_1 be I'_1, as shown in Fig. 11.1.

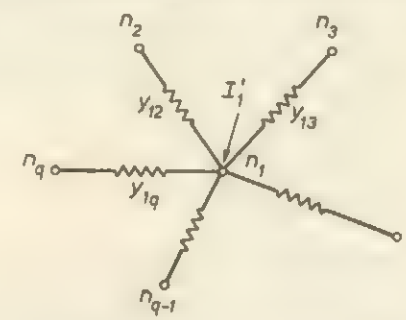


FIG. 11.1. COMPUTING THE NODE-PAIR MATRIX BY INSPECTION

The admittances connected to n_1 are $y_{12}, \ldots, y_{1q}$, and the potentials of $n_1, \ldots, n_q$ relative to some reference node are $E'_1, \ldots, E'_q$. It follows from Ohm's law that

$$\begin{aligned} I'_1 &= (E'_1 - E'_2)y_{12} + \ldots + (E'_1 - E'_q)y_{1q} \\ &= (y_{12} + \ldots + y_{1q})E'_1 - y_{12}E'_2 - \ldots - y_{1q}E'_q \qquad (11.3) \end{aligned}$$

This equation enables us to write down $\mathbf{Y}'$ by inspection according to the following rules—

(i) The diagonal element in position kk of $\mathbf{Y}'$ is equal to the sum of the admittances of the branches converging on node k.

(ii) The element in position km ($k \neq m$) is equal to the negative of the admittance y_{km} ($= y_{mk}$) joining nodes k and m.

For the subsequent development of the theory of diakoptics, it is of vital importance to note carefully what happens to $\mathbf{Y}'$ if the branch connecting nodes k and m is removed.

This change will affect the elements in positions kk, km, mm and mk as follows: The elements y'_{kk} and y'_{mm} will both be reduced by

the amount $y_{km} = y_{mk}$, and the elements y_{km} and y_{mk} in positions km and mk will be deleted.

To conclude this section, we give in Table 11.1 a list of symbols and nomenclature to be employed subsequently.

Table 11.1

NOTATION AND NOMENCLATURE

Symbol	Name and Definition
$\mathbf{Y}$	(system) admittance matrix = primitive admittance matrix of the network to be analysed by diakoptics.
$\mathbf{Y}'$	(system) node-pair admittance matrix (defined in Section 10.4).
$\mathbf{Y}''$	(system) torn admittance matrix = node-pair admittance matrix of the unconnected network that remains when the tie branches are removed.
$\mathbf{Z}''(= \mathbf{Y}''^{-1})$	torn impedance matrix.
$\mathbf{z}(= \mathbf{y}^{-1})$	tie impedance matrix = diagonal matrix containing all the tie branch impedances.
$\mathbf{K}$	tie connexion matrix = matrix indicating how the tie branches are connected to the nodes of the original network.

11.2. Division of Network into Subnetworks

In this section we shall discuss what happens to the node-pair matrix $\mathbf{Y}'$ of a large system when a number of branches are removed

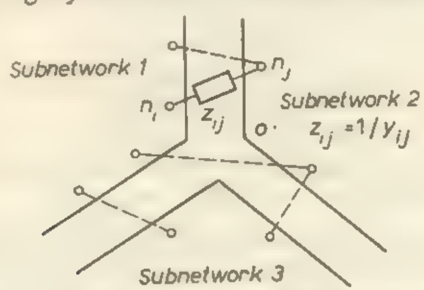


FIG. 11.2. DIVISION OF NETWORK INTO SUBNETWORKS

so as to divide the network into a number of uncoupled subnetworks.

We shall assume that the non-torn network is connected and that the p independent node potentials are measured relative to some datum node which is common to all the subnetworks. Kron calls such a network a *diffusion-type network*.

The tie immittances, typified by the impedance z_{ij} $(= 1/y_{ij})$, are assumed to have no magnetic couplings with any of the subnetworks 1, 2, ...

The p-by-p node-pair matrix $\mathbf{Y}'$ is now partitioned in accordance with the division of the network as follows—

$$\mathbf{Y}' = \begin{Bmatrix} \mathbf{Y}'_{11} & \mathbf{Y}'_{12} & \mathbf{Y}'_{13} \\ \mathbf{Y}'_{21} & \mathbf{Y}'_{22} & \mathbf{Y}'_{23} \\ \mathbf{Y}'_{31} & \mathbf{Y}'_{32} & \mathbf{Y}'_{33} \end{Bmatrix}$$

All the submatrices along the main matrix diagonal are square, and their off-diagonal elements depend only on the corresponding subnetwork.

The elements of the non-diagonal submatrices $\mathbf{Y}'_{ij}$ $(i \neq j)$ consist of the tie admittances connecting the nodes of subnetworks i and j. When one or more of the subnetworks contain active devices such as vacuum tubes, the corresponding matrices on the main matrix diagonal of $\mathbf{Y}'$ will be non-symmetric, but, because of our choice of ties, the remaining portion of $\mathbf{Y}'$ will be symmetric, i.e. $\mathbf{Y}'_{ij} = (\mathbf{Y}'_{ji})_t$ $(i \neq j)$.

If we now remove all the tie admittances, $\mathbf{Y}'$ will be transformed into $\mathbf{Y}''$ according to the pattern outlined in the previous section: all the submatrices not on the main matrix diagonal vanish, and all the elements of row (or column) i which do not belong to the main matrix diagonal are added to the elements in position ii.

To compensate for the removal of the tie branch connecting nodes i and j and oriented from (say) i to j, we inject currents $-I''$ and $+I''$ at nodes i and j, respectively, where I'' is equal to the current flowing in the branch connecting node i with node j.

These injected currents are added to the externally injected currents, I'_i and I'_j. In order to do so, we compile a p-by-r *tie connexion matrix* $\mathbf{K}$ (where r is the number of tie branches) with a row for each tie branch. In column q of $\mathbf{K}$, $+1$ will appear in the row corresponding to the node where $+I''_q$ is injected, and -1 in the row of the node at which $-I''_q$ is applied.

$$\begin{array}{c} r \text{ columns} \\ \mathbf{K} = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & -1 & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & +1 & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \begin{array}{l} \\ i \\ \; p \text{ rows} \\ j \\ \\ \end{array} \\ q \end{array}$$

Thus the matrix product $\mathbf{K}I''$ orders the $2r$ tie branch currents according to the nodes at which they are injected. The augmented set of injected currents of the torn network, arranged in the usual way as a p-dimensional vector, is

$$I_1' = I' + \mathbf{K}I'' \quad . \quad . \quad . \quad . \quad (11.4)$$

where I' represents the original injected node-pair currents, and the r-dimensional vector I'' represents the currents that are injected to compensate for the removal of the tie branches.

From the manner in which $\mathbf{K}$ was defined, each of its columns indicates and orients the nodes forming the terminals of a tie branch. It is readily seen that the p.d.s across the r tie branches can be expressed as $-\mathbf{K}_t E'$ when the orientation of currents and voltage drops is chosen in accordance with the polarity convention given in Fig. 10.1, p. 112.

Finally, arranging the r tie impedances z_{ij} along the main diagonal of an r-by-r matrix $\mathbf{z}$, we have

$$\mathbf{K}_t E' = -\mathbf{z}I'' \quad . \quad . \quad . \quad . \quad (11.5)$$

11.3. Inversion of Node-Pair Matrix $\mathbf{Y}'$

It was shown in the previous section how the removal of the tie immittances was compensated for by injecting a set of currents $\mathbf{K}I''$ at the node pairs of the torn network in addition to the set I' which were present in the original network.

The node-pair equation of the torn network is

$$I' + \mathbf{K}I'' = \mathbf{Y}''E' \quad . \quad . \quad . \quad (11.6)$$

or

$$E' = (\mathbf{Y}'')^{-1}(I' + \mathbf{K}I'') = \mathbf{Z}''(I' + \mathbf{K}I'') \quad . \quad . \quad (11.7)$$

This equation is not, however, a true solution to the network problem, since the node-pair currents I'' which appear on its right-hand side are in reality branch currents and, as such, unknown response quantities. To eliminate I'', we utilize eqn. (11.5). Pre-multiplication of eqn. (11.7) by $\mathbf{K}_t$ gives

$$\mathbf{K}_t E' = \mathbf{K}_t \mathbf{Z}''(I' + \mathbf{K}I'') = -\mathbf{z}I'' \quad . \quad . \quad (11.8)$$

which, solved for I'', yields

$$I'' = -(\mathbf{z} + \mathbf{K}_t\mathbf{Z}''\mathbf{K})^{-1}\mathbf{K}_t\mathbf{Z}''I' \quad . \quad . \quad (11.9)$$

Substituting this equation in eqn. (11.7), we find

$$E' = \mathbf{Z}''(\mathbf{I} - \mathbf{K}(\mathbf{z} + \mathbf{K}_t\mathbf{Z}''\mathbf{K})^{-1}\mathbf{K}_t\mathbf{Z}'')I' \quad . \quad (11.10)$$

which is a solution to the original network problem

$$I' = \mathbf{Y}'E' \quad . \quad . \quad . \quad . \quad (11.11)$$

By comparison of eqns. (11.10) and (11.11), it is clear that

$$(\mathbf{Y}')^{-1} = \mathbf{Z}'' - \mathbf{Z}''\mathbf{K}(\mathbf{z} + \mathbf{K}_t\mathbf{Z}''\mathbf{K})^{-1}\mathbf{K}_t\mathbf{Z}'' \quad . \quad (11.12)$$

The main advantage of diakoptics can be read from eqn. (11.10). The only matrix inversions to be performed in order to solve eqn. (11.11) are

$$(\mathbf{Y}'')^{-1} = \begin{bmatrix} \mathbf{Y}_{11}'' & \mathbf{0} & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{22}'' & \cdot & \cdot & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \cdot & \cdot & \mathbf{Y}_{qq}'' \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} (\mathbf{Y}_{11}'')^{-1} & \mathbf{0} & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & (\mathbf{Y}_{22}'')^{-1} & \cdot & \cdot & \mathbf{0} \\ \cdot & \mathbf{0} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \mathbf{0} & \cdot & \cdot & (\mathbf{Y}_{qq}'')^{-1} \end{bmatrix} = \mathbf{Z}''$$

which is comparatively simple, and

$$(\mathbf{z} + \mathbf{K}_t\mathbf{Z}''\mathbf{K})^{-1}$$

which is an r-by-r matrix where $r < p$.

Exercise 1. In addition to the two above-mentioned inversions, we may also invert the diagonal r-by-r tie admittance matrix $\mathbf{y}$ ($\mathbf{z} = \mathbf{y}^{-1}$), although this is not strictly necessary. Pre-multiplying the term $\mathbf{K}_t\mathbf{Z}''\mathbf{K}$ by $\mathbf{zy}$ ($= \mathbf{I}$), we get

$$(\mathbf{z} + \mathbf{zy}\mathbf{K}_t\mathbf{Z}''\mathbf{K})^{-1} = (\mathbf{I} + \mathbf{y}\mathbf{K}_t\mathbf{Z}''\mathbf{K})^{-1}\mathbf{z}^{-1} = (\mathbf{I} + \mathbf{y}\mathbf{K}_t\mathbf{Z}''\mathbf{K})^{-1}\mathbf{y}$$

Similarly, post-multiplication of $\mathbf{K}_t\mathbf{Z}''\mathbf{K}$ by $\mathbf{yz}$ ($= \mathbf{I}$) will yield

$$(\mathbf{z} + \mathbf{K}_t\mathbf{Z}''\mathbf{K})^{-1} = \mathbf{y}(\mathbf{I} + \mathbf{K}_t\mathbf{Z}''\mathbf{K}\mathbf{y})^{-1}$$

neither of which contains $\mathbf{z}$.

Eqn. (11.12) is so important to the theory of diakoptics that we shall derive it once more in another way.

Consider the product $\mathbf{K}\mathbf{y}\mathbf{K}_t$, where $\mathbf{y}$ is the r-by-r diagonal tie admittance matrix, and $\mathbf{K}$ the p-by-r tie connexion matrix. It will be recalled from Section 11.2 that each column of $\mathbf{K}$ indicates and orients the two nodes between which the corresponding tie branch is connected.

Take element y_{qq} of $\mathbf{y}$ and let us assume that this tie connects nodes u and v of the original network. Ignoring for the time being all other elements of $\mathbf{y}$, the product $\mathbf{Ky}$ will be

$$\mathbf{Ky} = \begin{array}{c} \\ u \\ \\ \\ v \\ \\ \end{array}\!\!\begin{array}{c} q \\ \left\{\begin{matrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & +1 & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & -1 & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{matrix}\right\} \end{array} \begin{array}{c} q \\ \left\{\begin{matrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & y_{qq} & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{matrix}\right\} \end{array}\!\!\begin{array}{c} \\ \\ q \\ \\ \end{array}$$

$$= \begin{array}{c} q \\ \left\{\begin{matrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & +y_{qq} & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & -y_{qq} & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{matrix}\right\} \end{array}\!\!\begin{array}{c} \\ \\ u \\ \\ \\ v \\ \\ \end{array}$$

Post-multiplication of this expression by $\mathbf{K}_t$ yields a p-by-p matrix

$$\mathbf{KyK}_t = \begin{array}{c} \quad u \qquad\qquad\quad v \\ \left\{\begin{matrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & +y_{qq} & \cdot & \cdot & -y_{qq} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & -y_{qq} & \cdot & \cdot & +y_{qq} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}\right\} \end{array}\!\!\begin{array}{c} \\ \\ u \\ \\ \\ v \\ \\ \end{array}$$

which has $+y_{qq}$ in positions uu and vv, and $-y_{qq}$ in positions uv and vu.

This result is readily generalized to the case where all the diagonal elements of $\mathbf{y}$ are non-zero. By comparison with the system outlined in Section 11.1 for writing down $\mathbf{Y}'$ directly from inspection, it is clear that

$$\mathbf{Y}' = \mathbf{Y}'' + \mathbf{KyK}_t \quad . \quad . \quad . \quad (11.13)$$

Comparing this decomposition of $\mathbf{Y}'$ with eqns. (4.25) and (4.31), we again arrive at the expression for the inverse of $\mathbf{Y}'$.

Exercise 2. Yet another way of deriving eqn. (11.12) can be found by following up a hint given in Section 5.5, where the partitioned inverse of a matrix was derived.

Let a partitioned matrix

$$\begin{Bmatrix} -\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{Bmatrix} \text{ and its inverse } \begin{Bmatrix} \mathbf{K} & \mathbf{L} \\ \mathbf{M} & \mathbf{N} \end{Bmatrix}$$

be given, where $\mathbf{A}$, $\mathbf{D}$, $\mathbf{K}$ and $\mathbf{N}$ are square and the submatrices in corresponding positions are like.

Since inverses commute, we have

$$\begin{Bmatrix} -\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{Bmatrix}\begin{Bmatrix} \mathbf{K} & \mathbf{L} \\ \mathbf{M} & \mathbf{N} \end{Bmatrix} = \begin{Bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{Bmatrix} \quad . \quad . \quad . \quad (11.14)$$

and

$$\begin{Bmatrix} \mathbf{K} & \mathbf{L} \\ \mathbf{M} & \mathbf{N} \end{Bmatrix}\begin{Bmatrix} -\mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{Bmatrix} = \begin{Bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{Bmatrix} \quad . \quad . \quad . \quad (11.15)$$

Hence, from eqn. (11.14),

$$-\mathbf{AL} + \mathbf{BN} = \mathbf{O} \quad \text{or} \quad \mathbf{L} = \mathbf{A}^{-1}\mathbf{BN} \quad . \quad . \quad (11.16)$$

and

$$\mathbf{CL} + \mathbf{DN} = \mathbf{I} \, . \quad . \quad . \quad . \quad . \quad (11.17)$$

Substituting eqn. (11.16) in eqn. (11.17),

$$\mathbf{N}^{-1} = \mathbf{D} + \mathbf{CA}^{-1}\mathbf{B} \quad . \quad . \quad . \quad . \quad (11.18)$$

Eqn. (11.15) yields

$$\mathbf{KB} + \mathbf{LD} = \mathbf{O} \quad \text{or} \quad \mathbf{L} = -\mathbf{KBD}^{-1} \quad . \quad . \quad (11.19)$$

and

$$-\mathbf{KA} + \mathbf{LC} = \mathbf{I} \quad . \quad . \quad . \quad . \quad (11.20)$$

Eliminating $\mathbf{L}$ from eqns. (11.19) and (11.20), we get

$$\mathbf{K} = -(\mathbf{A} + \mathbf{BD}^{-1}\mathbf{C})^{-1} \quad . \quad . \quad . \quad . \quad . \quad (11.21)$$

and, substituting this equation in eqn. (11.19),

$$\mathbf{L} = (\mathbf{A} + \mathbf{BD}^{-1}\mathbf{C})^{-1}\mathbf{BD}^{-1} \quad . \quad . \quad . \quad . \quad . \quad (11.22)$$

which, when substituted in eqn. (11.17), yields

$$\mathbf{N} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} + \mathbf{BD}^{-1}\mathbf{C})^{-1}\mathbf{BD}^{-1} \quad . \quad . \quad (11.23)$$

Substituting $(\mathbf{Y}')^{-1}$ for $\mathbf{N}$, $\mathbf{Y}''$ for $\mathbf{D}$, $\mathbf{Z}''$ for $\mathbf{D}^{-1}$, $\mathbf{K}$ for $\mathbf{C}$, $\mathbf{K}_t$ for $\mathbf{B}$, $\mathbf{y}$ for $\mathbf{A}^{-1}$, and $\mathbf{z}$ for $\mathbf{A}$, eqns. (11.18) and (11.23) transform into eqns. (11.13) and (11.12), respectively.

PROBLEMS

11.1. Verify the statement in Section 11.1 that the orientation of the individual branches has no effect on the sign of the elements of the mesh-impedance matrix $\mathbf{Z}'$.

11.2. Compute the node-pair connexion matrix $\mathbf{A}$ of the network shown in Fig. 11.3. Calculate its node-pair admittance matrix $\mathbf{Y}' = \mathbf{A}_t\mathbf{YA}$, and compare

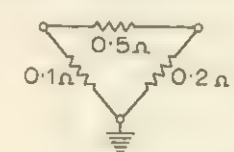


FIG. 11.3

the expression with that obtained by the method described in Section 11.1, i.e. by inspection.

11.3. Verify by detailed computation the statement in Exercise 1, Section 11.3, to the effect that the expression $\mathbf{Y}' = \mathbf{Y}'' + \mathbf{KyK}_t$ is generally valid.

11.4. Show that both $\mathbf{K}_t\mathbf{Z}''\mathbf{K}$ and $\mathbf{K}\mathbf{y}\mathbf{K}_t$ are invariant to any reorientation of the tie branches.

11.5. Write down the primitive admittance matrix $\mathbf{Y}$ of the network shown in Fig. 11.4 (all the admittances are given in millimhos) and compile the node-pair admittance matrix $\mathbf{Y}'$ by means of the formula $\mathbf{Y}' = \mathbf{A}_t\mathbf{Y}\mathbf{A}$, and also by inspection.

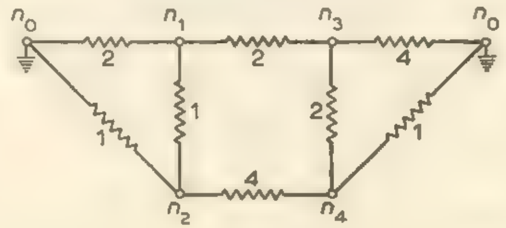


FIG. 11.4

"Tear" the network into two earthed subnetworks by removing branches $n_1 - n_3$ and $n_2 - n_4$, and calculate the torn admittance matrix $\mathbf{Y}''$. Write down the tie connexion matrix $\mathbf{K}$, and the tie impedance matrix $\mathbf{z}$. Calculate $(\mathbf{Y}')^{-1}$ by means of eqn. (11.12), and check the result by calculating $(\mathbf{Y}')(\mathbf{Y}')^{-1}$.

CHAPTER 12

Tensor Analysis

12.1. Introduction—Non-Linear Coordinate Transformations

It is customary for the author of a textbook to write a preface in which he apologizes for "adding yet another book to the already extensive literature on the subject." He then goes on to explain why he considers the publication of his work to be justified, and points to a novel method of exposition or some other feature in order to prove his contention.

Having thus excused himself, thanked his colleagues and his publishers, and prepared his reader for what is to follow, the author proceeds to present his subject only to come to an abrupt stop at the end of the final chapter. Since even the largest textbook cannot hope to explore fully all the aspects of any given subject, this sudden conclusion leaves the reader suspended in mid-air not knowing what to do next. To match his polite introduction it would be fitting if the author were to take his leave by pointing out some of the things which lack of space had compelled him to ignore, and to suggest a further course of study for those interested readers who might desire to learn more about the subject.

It was mentioned in Chapter 1 that we would endeavour to present the theory of matrices so as to prepare the way for the theory of tensors. It would therefore be natural to conclude the book with a brief account of the elements of tensor analysis. The present chapter should thus be viewed, not as an introduction to tensor, calculus, but rather as an appendix to a book on matrices which points to the path any student must follow if he wishes to penetrate deeper into the world of the geometrized physical sciences, where matrix algebra, tensor analysis, group theory and topology combine with physics and engineering to form a higher unity.

A variable point P in n-dimensional space (V_n) is expressed in a coordinate system S_x by a set of n ordered variables x^i ($i = 1, \ldots, n$). Relative to some other system S_y, the same point P is given by another set of variables y^i. The coordinate transformation from S_x to S_y is said to be *admissible* in a region R, if the $y^i = y^i(x^j)$ are differentiable single-valued functions of the x^j such that the Jacobian determinant $\det(\partial y^i/\partial x^j) \neq 0$ in R.

An admissible transformation is invertible, the x^i being differentiable single-valued functions of the y^i, and the corresponding Jacobian is

$$\det(\partial x^i/\partial y^j) = 1/\det(\partial y^i/\partial x^j) \qquad (i, j = 1, \ldots, n) \quad (12.1)$$

Also, since both sets of variables are independent, we have

$$\frac{\partial x^i}{\partial x^j} = \frac{\partial y^i}{\partial y^j} = \delta^i_j = \begin{Bmatrix} 1 \text{ when } i = j \\ 0 \text{ when } i \neq j \end{Bmatrix} \quad . \quad . \quad (12.2)$$

The symbol δ^i_j is the *Kronecker delta*, which, written as a matrix, is identical with I.

Applying a second transformation from S_y to S_z, we have $z^k = z^k(y^j(x^i))$, and hence, employing Einstein's summation convention,

$$\frac{\partial z^k}{\partial x^i} = \frac{\partial z^k}{\partial y^j} \cdot \frac{\partial y^j}{\partial x^i} \quad . \quad . \quad . \quad . \quad (12.3)$$

When $S_x = S_z$, and therefore $x^i = z^i$, this equation becomes

$$\frac{\partial x^k}{\partial x^i} = \delta^k_i = \frac{\partial x^k}{\partial y^j} \cdot \frac{\partial y^j}{\partial x^i} \quad . \quad . \quad . \quad . \quad (12.4)$$

It is obvious from the above equations that the set of partial derivatives corresponding to a set of admissible transformations form a group. The consecutive application of two admissible coordinate transformations is itself an admissible transformation (the combination of derivatives being given by eqn. (12.3); the identity transformation exists (with derivatives equal to the Kronecker delta); and every admissible transformation possesses an inverse transformation which is also admissible (the relevant derivatives being reciprocally related as shown in eqn. (12.4)).

When the y^i are linear homogeneous independent functions of the x^j, we have

$$y^i = a^i_j x^j \quad . \quad . \quad . \quad . \quad (12.5)$$

and hence

$$\partial y^i/\partial x^j = a^i_j \quad . \quad . \quad . \quad . \quad (12.6)$$

where $\det(a^i_j) \neq 0$. Compare these expressions with eqn. (6.5), p. 60.

An infinitesimal displacement of P is given as dx^i in S_x, and dy^i in S_y. The relation between the two differentials is found by differentiating the function $y^i = y^i(x^j)$. This yields

$$dy^i = \frac{\partial y^i}{\partial x^j} \cdot dx^j \quad . \quad . \quad . \quad . \quad (12.7)$$

To introduce the idea of distance into our space we define a quadratic form (a *metric*)

$$ds^2 \overset{\text{D}}{=} {}_x g_{ij} dx^i dx^j \quad . \quad . \quad . \quad . \quad (12.8)$$

where dx^i is the displacement of P, and ${}_x g_{ij}$, which is symmetric in i and j (*see* Exercise 1, Section 6.1), is the *metric* (or *fundamental*) tensor of our space referred to S_x.

Assuming distance to be invariant to coordinate transformations, we find, by using eqn. (12.7),

$$ds^2 = {}_x g_{ij} \frac{\partial x^i}{\partial y^r} \cdot \frac{\partial x^j}{\partial y^s} \cdot dy^r dy^s = {}_y g_{rs} dy^r dy^s \quad . \quad . \quad (12.9)$$

from which we deduce that the metric tensor transforms according to the formula

$${}_y g_{rs} = \frac{\partial x^i}{\partial y^r} \cdot \frac{\partial x^j}{\partial y^s} \, {}_x g_{ij} \quad . \quad . \quad . \quad (12.10)$$

12.2. Co- and Contravariant Tensors

A set of n functions ${}_x a^i$ of the n variables x^i are said to be the components of a *contravariant monovalent tensor* (*contravariant vector*) referred to the coordinate system S_x, if they transform according to the formula

$${}_y a^i = \frac{\partial y^i}{\partial x^j} \, {}_x a^j \quad . \quad . \quad . \quad . \quad (12.11)$$

when we apply a transformation belonging to a group of admissible coordinate transformations.

Analogously, a set of functions ${}_x a_i$ transforming according to the pattern

$${}_y a_i = \frac{x^j \partial}{\partial y^i} \, {}_x a_j \quad . \quad . \quad . \quad . \quad (12.12)$$

are the components in S_x of a *covariant vector* (*monovalent covariant tensor*).

Except in the case of Euclidean space referred to rectilinear coordinates, the coordinates themselves are not tensors (in spite of the fact that they are indicated by means of a kernel letter and a superscript).

From eqn. (12.7) it appears, however, that the dx^i are the components of a contravariant vector.

A simple example of a covariant vector can be found in the gradient of a scalar field $V = V(x^i)$. When transforming grad V from S_x to S_y, we have

$$\frac{\partial V}{\partial y^i} = \frac{\partial x^j}{\partial y^i} \cdot \frac{\partial V}{\partial x^j} \quad . \quad . \quad . \quad (12.13)$$

which is functionally identical with eqn. (12.12).

With the above equations of definition in mind, we now go on to define tensors of higher valence. The fundamental tensor, g_{ij}, which was introduced to metrize our space, is an example of a divalent covariant tensor with the transformation formulae given in eqn. (12.10) (*see* also Exercise 1, Section 6.1).

Similarly, a divalent contravariant tensor a^{ij} will obey the transformation equation

$${}_ya^{ij} = \frac{\partial y^i}{\partial x^r} \cdot \frac{\partial y^j}{\partial x^s}\, {}_xa^{rs} \quad . \quad . \quad . \quad (12.14)$$

A tensor with one contravariant and one covariant index is called a *divalent mixed* tensor. It transforms as follows—

$${}_ya^i_j = \frac{\partial y^i}{\partial x^r} \cdot \frac{\partial x^s}{\partial y^j}\, {}_xa^r_s \quad . \quad . \quad . \quad (12.15)$$

The Kronecker delta is an instance of such a mixed divalent tensor—

$${}_y\delta^i_j = \frac{\partial y^i}{\partial x^r} \cdot \frac{\partial x^s}{\partial y^j}\, {}_x\delta^r_s = \frac{\partial y^i}{\partial x^r} \cdot \frac{\partial x^r}{\partial y^j} = \frac{\partial y^i}{\partial y^j} = {}_x\delta^i_j \quad . \quad (12.16)$$

We have shown in Section 6.1, p. 60, that the distinction between co- and contravariance disappears when we confine ourselves to the group of orthonormal transformations. This result also applies to tensors but will not be proved here.

Note 1. To acquaint himself with the mechanism of indicial notation, the reader merely has to compare the various transformation formulae given up to now. By convention, the dx^i are termed contravariant, and superior and inferior indices have been chosen to denote contravariance and covariance respectively.

In tensor formulae the *live* or *free* (i.e. non-dummy) indices of the two members of an equation "check", the same index appearing in the same position on both sides of the equality sign. Also, it will be noted that we always sum over a pair of contragredient indices; we shall see later (Section 12.3) that this is necessary if we wish to preserve the tensor character of our expressions.

Different tensors are indicated by different kernel letters, and to remind us of the coordinate system to which a particular tensor is referred, we use an inferior prefix.

It will be noted from eqn. (12.13) that differentiation with respect to a contravariant variable results in an additional covariant index.

The reader must reconcile himself with the fact that it takes some time to get used to the kernel-index notation. With a little practice, however, he will discover its advantages and find out for himself what a powerful mnemonic aid the positioning of indices can be.

Exercise 1. If we define a divalent covariant tensor which is identical to the Kronecker delta in (say) S_x,

$${}_x\delta_{ij} \equiv {}_x\delta^i_j$$

we can investigate how its coordinates transform when we pass from S_x to S_y. We find that

$${}_y\delta_{ij} = \frac{\partial x^r}{\partial y^i} \cdot \frac{\partial x^s}{\partial y^j} \cdot {}_x\delta_{rs} = \frac{\partial x^r}{\partial y^i} \cdot \frac{\partial x^r}{\partial y^j} \neq {}_x\delta^{ij}$$

which proves that ${}_y\delta_{ij}$ is no longer a Kronecker delta. As an illustration of this consider the metric tensor of a 2-dimensional space. Referred to Cartesian coordinates, it is represented by the 2-by-2 unit matrix, whereas in polar coordinates (r, θ) it changes into

$$\begin{Bmatrix} 1 & 0 \\ 0 & r^2 \end{Bmatrix}$$

Exercise 2. We shall now test whether the property of symmetry of a divalent contravariant tensor a^{ij} is invariant to coordinate transformations.

By hypothesis, ${}_xa^{ij} = {}_xa^{ji}$, and from eqn. (12.14) we get

$${}_ya^{ij} = \frac{\partial y^i}{\partial x^r} \cdot \frac{\partial y^j}{\partial x^s} \cdot {}_xa^{rs} = \frac{\partial y^i}{\partial x^r} \cdot \frac{\partial y^j}{\partial x^s} \cdot {}_xa^{sr} = {}_ya^{ji}$$

which shows that the symmetry of a^{ij} is not destroyed by transformation. Symmetry is also preserved in a covariant divalent tensor, but the proof breaks down for a mixed divalent tensor which happens to be symmetric with respect to its contragredient indices in a specific coordinate system.

The transformation equations given in this section can easily be generalized to cater for tensors of higher valence than two. For instance, a^k_{ij} is a trivalent tensor, contravariant in one index and covariant in two. Its equation of transformation is

$${}_ya^k_{ij} = \frac{\partial x^r}{\partial y^i} \cdot \frac{\partial x^s}{\partial y^j} \cdot \frac{\partial y^k}{\partial x^t} \cdot {}_xa^t_{rs}$$

Note 2. In any given coordinate system the components of a tensor can be written in the form of a matrix (in the case of tensors of valence higher than 2 a "matrix of matrices" will be required). It should be clearly understood that a matrix as such is not a tensor. A matrix can be given the name *tensor* only when the underlying transformation group and the equation of transformation are known.

It is quite possible for a set of functions to be the components of a tensor under one group of transformations but not under another more general group. Hence, only the orthogonal form of the matrix equations of an electrical network can be said to be tensor equations, the associated group being the

connexion group, the elements of which are the square mesh-node connexion matrices corresponding to all the networks that can be built from the given number of branches.

12.3. Tensor Algebra

We can easily convince ourselves, by reference to the defining equations of the previous section, that any linear combination of tensors of the same type is a tensor of the same type, provided that the constants appearing as coefficients are invariant to the relevant group of transformations. An important property of tensors implicit in this statement is that, if a tensor vanishes in one coordinate system, it will vanish in all other coordinate systems of the underlying group.

From two arbitrary tensors we can form another tensor, the *outer product*, whose valence is the sum of those of its constituent factors (*see* Section 5.1, p. 51). Thus, for instance,

$$a_i^j b_{km} = c_{ikm}^j$$

Also, from a mixed tensor of valence N we can form a tensor of valence $N - 2$ by *contraction*, i.e. by summing over a pair of contragredient indices. Hence, the mixed trivalent tensor a_k^{ij} becomes a contravariant vector when we perform the summation a_i^{ij}. Contraction does not alter the tensor character of a_k^{ij}: from the transformation equation

$$_y a_k^{ij} = \frac{\partial y^i}{\partial x^r} \cdot \frac{\partial y^j}{\partial x^s} \cdot \frac{\partial x^t}{\partial y^k} \cdot {}_x a_t^{rs} \quad . \quad . \quad . \quad (12.17)$$

we get

$$_y a_i^{ij} = \frac{\partial y^i}{\partial x^r} \cdot \frac{\partial y^j}{\partial x^s} \cdot \frac{\partial x^t}{\partial y^i} \cdot {}_x a_t^{rs}$$

$$= \frac{\partial y^j}{\partial x^s} \cdot \delta_r^t \cdot {}_x a_t^{rs} = \frac{\partial y^j}{\partial x^s} \cdot {}_x a_r^{rs} \quad . \quad . \quad (12.18)$$

which clearly demonstrates that a_i^{ij} is a contravariant vector.

The contraction of a tensor with respect to a pair of cogredient indices is also possible, but this operation does not yield a tensor—a fact which is easily verified.

Contracting a divalent mixed tensor, we obtain an invariant scalar quantity (the *trace*) (cf. Section 6.7, p. 73). It is thus consistent with our terminology to call an invariant a *tensor of zero valence*.

We have proved that the outer product of two arbitrary tensors is a tensor. Conversely, it can be proved that, if the outer product of a set of functions by an arbitrary tensor is a tensor, the set of functions are the components of a tensor of the type necessary to

balance the equation indicially. This theorem is called the *quotient law* (we shall not prove it here).

By employing outer multiplication followed by contraction, we now introduce a contravariant divalent symmetric tensor g^{ij}, the contravariant metric tensor, which satisfies the equation

$$g_{ij} g^{jk} = \delta_i^k \quad . \quad . \quad . \quad (12.19)$$

In any coordinate system g_{ij} and g^{mn} are represented by reciprocal matrices.

Outer multiplication coupled with contraction is termed *inner multiplication.* Inner multiplication of a contravariant vector by the covariant tensor results in a covariant vector

$$a^r g_{rs} = a_s \quad . \quad . \quad . \quad . \quad (12.20)$$

This process is called *lowering an index.* The reverse process of *raising an index* is accomplished by inner multiplication by the contravariant metric tensor g^{ij}.

Exercise. A tensor can be contracted with respect to two cogredient indices, but the transformation equation indicates that the result is not a tensor. Thus, for instance,

$$_x a^{ii} = \frac{\partial x^i}{\partial y^j} \cdot \frac{\partial x^i}{\partial y^k} \cdot {}_y a^{jk}$$

which is not equal to $_y a^{kk}$, because $\dfrac{\partial x^i}{\partial y^j} \cdot \dfrac{\partial x^i}{\partial y^k}$ is not generally identical with the Kronecker delta.

12.4. The Christoffel Symbols

In general, the components of the metric tensors are functions of the space coordinates. We shall now define two sets of derived functions, which, as we shall see, do not form the components of a tensor but are nevertheless necessary in order to develop the concepts of covariant and intrinsic differentiation.

The *Christoffel symbol of the first kind* is defined by the equation

$$[ij, k] \overset{\text{D}}{=} \frac{1}{2} \left[\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k} \right] \quad . \quad . \quad (12.21)$$

The *Christoffel symbol of the second kind* is defined by

$$\left\{ \begin{matrix} m \\ i \quad j \end{matrix} \right\} \overset{\text{D}}{=} [ij, k]\, g^{km} \quad . \quad . \quad . \quad (12.22)$$

In order to obtain the transformation formula of $[ij, k]$, we differentiate eqn. (12.10) with respect to y^k and find

$$\frac{\partial\, {}_y g_{ij}}{\partial y^k} = \frac{\partial\, {}_x g_{rs}}{\partial x^t} \cdot \frac{\partial x^t}{\partial y^k} \cdot \frac{\partial x^r}{\partial y^i} \cdot \frac{\partial x^s}{\partial y^j} + \frac{\partial^2 x^r}{\partial y^i \partial y^k} \cdot \frac{\partial x^s}{\partial y^j} \cdot {}_x g_{rs} + \frac{\partial x^r}{\partial y^i} \cdot \frac{\partial^2 x^s}{\partial y^j \partial y^k} \cdot {}_x g_{rs} \quad . \quad . \quad . \quad (12.23)$$

which shows, incidentally, that $\partial_y g_{ij}/\partial y^k$ is not a tensor unless the second derivatives vanish identically.

Permuting the indices in this equation, and using eqn. (12.21), we find, after some algebraic manipulation,

$${}_y[ij, k] = {}_x[rs, t] \cdot \frac{\partial x^r}{\partial y^i} \cdot \frac{\partial x^s}{\partial y^j} \cdot \frac{\partial x^t}{\partial y^k} + \frac{\partial x^r}{\partial y^k} \cdot \frac{\partial^2 x^s}{\partial y^i \partial y^j} \cdot {}_x g_{rs} \quad (12.24)$$

Multiplying the members of this equation by the corresponding members of the transformation equation

$${}_y g^{kn} = \frac{\partial y^k}{\partial x^t} \cdot \frac{\partial y^n}{\partial x^u} \cdot {}_x g^{tu} \quad . \quad . \quad . \quad (12.25)$$

we arrive at the expression

$${}_y\left\{\begin{matrix} n \\ i \quad j \end{matrix}\right\} = {}_x\left\{\begin{matrix} u \\ r \quad s \end{matrix}\right\} \frac{\partial y^n}{\partial x^u} \cdot \frac{\partial x^r}{\partial y^i} \cdot \frac{\partial x^s}{\partial y^j} + \frac{\partial y^n}{\partial x^s} \cdot \frac{\partial^2 x^s}{\partial y^i \partial y^j} \quad . \quad (12.26)$$

Eqns. (12.24) and (12.26) clearly indicate that neither of the Christoffel symbols is a tensor.

Inner multiplication of eqn. (12.26) by $\partial x^r/\partial y^n$ yields the identity

$$\frac{\partial^2 x^t}{\partial y^i \partial y^j} = {}_y\left\{\begin{matrix} n \\ i \quad j \end{matrix}\right\} \frac{\partial x^t}{\partial y^n} - {}_x\left\{\begin{matrix} t \\ r \quad s \end{matrix}\right\} \frac{\partial x^r}{\partial y^i} \cdot \frac{\partial x^s}{\partial y^j} \quad . \quad (12.27)$$

expressing the second derivative of x^t with respect to y^i in terms of the Christoffel symbols and the first derivatives.

Exercise. From the defining equations it is clear that the Christoffel symbols are symmetric with respect to the indices i and j. Hence, by interchanging i and k in eqn. (12.21) and adding the resulting equations, we get

$$\frac{\partial g_{ik}}{\partial x^j} = [ij, k] + [ik, j] \quad . \quad . \quad . \quad . \quad (12.28)$$

Note. At first sight one might be tempted to believe that the Christoffel symbols, being derived from the fundamental tensor, would also possess tensor characteristics. The existence of a second term in their transformation equations proves that such a supposition is untrue. This points to the significant fact that a tensor equation is invariant in functional form with respect to its associated transformation group.

It follows that if we succeed in expressing a physical relationship as a tensor equation, it will acquire the status of a "natural law" by virtue of its formal invariance. Under the circumstances, we ought to be justified in assuming the physical concepts entering into the equation, as well as the "law" itself, to be of fundamental importance seeing that they do not depend on the system of coordinates (within the specified group) to which they are referred.

Kron's *second generalization postulate* states this idea in other words. Observe in this connexion the functional identity between the orthogonal network equation and Ohm's law for a simple series circuit containing a source of e.m.f. e and an impedance $Z : e = Zi$.

12.5. Covariant Differentiation

Eqn. (12.23) indicated that the derivative of a tensor need not itself be a tensor. The question, therefore, naturally arises, Can we define a system of differentiation such that the derivative of a tensor will still be a tensor? This section will provide an affirmative answer.

Let us begin by differentiating the transformation equation of a contravariant vector (eqn. (12.11)) with respect to the variable y^k—

$$\frac{\partial\, {}_y a^i}{\partial y^k} = \frac{\partial\, {}_x a^j}{\partial x^m} \cdot \frac{\partial x^m}{\partial y^k} \cdot \frac{\partial y^i}{\partial x^j} + {}_x a^j \cdot \frac{\partial^2 y^i}{\partial x^j \partial x^m} \cdot \frac{\partial x^m}{\partial y^k} \quad . \quad (12.29)$$

Eliminating the second derivative by means of eqn. (12.27), we find

$$\frac{\partial\, {}_y a^i}{\partial y^k} = \frac{\partial\, {}_x a^j}{\partial x^m} \cdot \frac{\partial x^m}{\partial y^k} \cdot \frac{\partial y^i}{\partial x^j} + {}_x a^j \cdot \frac{\partial x^m}{\partial y^k}\, {}_x\left\{\begin{matrix} n \\ j \quad m \end{matrix}\right\} \frac{\partial y^i}{\partial x^n}$$
$$- {}_x a^j\, {}_y\left\{\begin{matrix} i \\ r \quad s \end{matrix}\right\} \frac{\partial y^r}{\partial x^j} \cdot \frac{\partial y^s}{\partial x^m} \cdot \frac{\partial x^m}{\partial y^k}. \quad . \quad (12.30)$$

from which, with the help of eqn. (12.11) and by changing a few dummy indices, we derive

$$\frac{\partial\, {}_y a^i}{\partial y^k} + {}_y a^r \left\{\begin{matrix} i \\ r \quad k \end{matrix}\right\} = \left[\frac{\partial\, {}_x a^n}{\partial x^m} + {}_x a^j \left\{\begin{matrix} n \\ j \quad m \end{matrix}\right\}\right] \frac{\partial x^m}{\partial y^k} \cdot \frac{\partial y^i}{\partial x^n} \quad (12.31)$$

This equation shows that the left-hand member transforms as a mixed divalent tensor (note the invariance in functional form).

The operation, which is indicated by means of a comma, is called *covariant differentiation*. Thus—

$${}_y a^i_{,k} \overset{D}{=} \frac{\partial\, {}_y a^i}{\partial y^k} + {}_y a^r \left\{\begin{matrix} i \\ r \quad k \end{matrix}\right\} \quad . \quad . \quad (12.32)$$

Under the group of linear transformations, the Christoffel symbols vanish identically, and covariant differentiation degenerates into ordinary partial differentiation.

By a train of reasoning analogous to that employed above, we can prove that the following covariant derivative of a covariant tensor—

$$ {}_x a_{i,k} \overset{\text{D}}{=} \frac{\partial_x a_i}{\partial x^k} - {}_x a_j \begin{Bmatrix} j \\ i \quad k \end{Bmatrix}_x \qquad . \quad . \quad . \; (12.33)$$

is also a tensor.

Similarly, a mixed tensor possesses a covariant derivative with tensor properties—

$$ {}_x a^i_{j,k} \overset{\text{D}}{=} \frac{\partial_x a^i_j}{\partial x^k} + {}_x a^i_m \begin{Bmatrix} m \\ j \quad k \end{Bmatrix}_x - {}_x a^n_j \begin{Bmatrix} i \\ n \quad k \end{Bmatrix}_x \qquad . \; (12.34)$$

This defining equation is readily generalized to tensors of any valence and any type.

Exercise. Applying covariant differentiation to the Kronecker delta

$$\begin{aligned} \delta^i_{j,k} &= \frac{\partial \delta^i_j}{\partial x^k} + \delta^n_j \begin{Bmatrix} i \\ n \quad k \end{Bmatrix} - \delta^i_m \begin{Bmatrix} m \\ j \quad k \end{Bmatrix} \\ &= 0 + \begin{Bmatrix} i \\ j \quad k \end{Bmatrix} - \begin{Bmatrix} i \\ j \quad k \end{Bmatrix} = 0 \end{aligned}$$

we see that the Kronecker delta behaves as a constant with respect to covariant differentiation

Also,

$$\begin{aligned} g_{ij,k} &= \frac{\partial g_{ij}}{\partial x^k} - g_{nj} \begin{Bmatrix} n \\ i \quad k \end{Bmatrix} - g_{im} \begin{Bmatrix} m \\ j \quad k \end{Bmatrix} \\ &= \frac{\partial g_{ij}}{\partial x^k} - [ik, j] - [jk, i] \\ &= 0 \qquad \text{by eqn. (12.28)} \end{aligned}$$

Likewise, we can prove that $g^{ij}_{,k} = 0$.

Covariant differentiation adds a covariant index to a tensor thus increasing its valence by one. Further, it can be shown that covariant differentiation obeys the rules of ordinary differentiation, the fundamental tensors and the Kronecker delta behaving like constants. Thus, for instance,

$$\begin{aligned} (a^i + b^i)_{,j} &= a^i_{,j} + b^i_{,j} \\ (a^i_j b^k)_{,m} &= a^i_{j,m} b^k + a^i_j b^k_{,m} \end{aligned}$$

and

$$(g^{ij} a_i)_{,k} = g^{ij} a_{i,k}$$

12.6. Intrinsic Differentiation

Differentiation of a tensor with respect to a scalar parameter t can be defined in such a way as to preserve the tensor character of the derivative, by means of the following equation—

$$\frac{\delta_x a^i_j}{\delta t} \overset{\text{D}}{=} {}_x a^i_{j,k} \cdot \frac{dx^k}{dt} \qquad . \quad . \quad . \; (12.35)$$

This derivative, called the *intrinsic derivative*, is a tensor, since it is formed as the inner product of the two tensors $a^i_{j,k}$ and dx^k/dt.

When the Christoffel symbols vanish, the intrinsic derivative becomes the ordinary scalar derivative.

The intrinsic derivative of an invariant is equal to the ordinary scalar derivative—

$$\frac{\delta a}{\delta t} = a_{,k} \frac{dx^k}{dt} = \frac{\partial a}{\partial x^k} \cdot \frac{dx^k}{dt} = \frac{da}{dt} \qquad . \quad . \; (12.36)$$

From the results of the Exercise in Section 12.5, it is clear that the intrinsic derivatives of the fundamental tensors and of the Kronecker delta all vanish identically.

Exercise. In polar coordinates, the components of the Christoffel symbol of the second kind are

$$\begin{Bmatrix} k \\ i \quad j \end{Bmatrix} = \left\{ \begin{array}{c} k \longrightarrow \\ \begin{array}{cc} & j \rightarrow \\ i \downarrow & \begin{Bmatrix} 0 & 0 \\ 0 & -r \end{Bmatrix} \end{array} \begin{array}{cc} & j \rightarrow \\ i \downarrow & \begin{Bmatrix} 0 & 1/r \\ 1/r & 0 \end{Bmatrix} \end{array} \end{array} \right\}$$

where we have written the symbol out in full as a 1-by-2 matrix of 2-by-2 matrices.

The components of a velocity are dx^i/dt ($i = 1, 2$) and dy^i/dt ($y^1 = r, y^2 = \theta$), in Cartesian and polar coordinates respectively. To obtain the components of the acceleration, we cannot differentiate the components of the velocity vector in the usual way, since this would not generally yield a vector quantity, but must resort to intrinsic differentiation.

In the Cartesian case, the Christoffel symbols vanish and we get d^2x^i/dt^2 as the components of the acceleration vector.

For a polar system, however, we have

$$\frac{\delta}{\delta t}\left(\frac{dy^i}{dt}\right) = \frac{d^2 y^i}{dt^2} + \frac{dy^j}{dt} \cdot \begin{Bmatrix} i \\ j \quad k \end{Bmatrix} \cdot \frac{dy^k}{dt}$$

which, after some manipulation, yields the components

$$\frac{d^2 r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 \quad \text{and} \quad \frac{d^2\theta}{dt^2} + \frac{2}{r} \cdot \frac{dr}{dt} \cdot \frac{d\theta}{dt}$$

(a well-known result).

To convince himself that these expressions are in fact correct, the reader would be well advised to transform the components d^2x^i/dt^2 by means of the transformation formula $(\partial y^i/\partial x^j)(d^2x^j/dt^2)$.

12.7. Conclusion and Prospect

And so we come to the end of the book, but not by any means to the end of the subject.

A firm grasp of the essentials of matrix algebra having been acquired, the logical step forward is to study the three interconnected disciplines: tensor analysis, group theory and algebraic topology.

This chapter has already given us a brief introduction to tensor calculus. We have defined tensors of various types, and generalized

the process of differentiation to enable it to take into account, not only the change in the coordinates of an object, but also the twisting and turning of the coordinates axes themselves, thus preserving the functional invariance which is such an important characteristic of tensors.

The immediate goal of the student should be to understand how a space is defined by means of its transformation group together with its metric tensor. Having metrized our space by defining the length of an infinitesimal displacement, our next step is to study the idea of the shortest distance between two points. Such geodesic problems appear in numerous disguises in many fields of applied mathematics, including the theory of electro-mechanical systems.

From this beginning the concept of curvature follows naturally. By means of certain curvature tensors, some of which are tetravalent, it becomes possible to detect the curvature of a given space "from the inside."

A large number of books are available on tensor analysis, those by McConnell,[5] Spain[8] and Sokolnikoff[7] providing comprehensive treatments which should suffice for most of the technical literature employing tensors. For more advanced texts the student may consult the books by Schouten.[9, 10]

In attempting a general description of the theory of groups, one can do no better than to quote Alice's remarks after seeing the Cheshire cat gradually fade away: "Well! I've often seen a cat without a grin, but a grin without a cat! It's the most curious thing I ever saw in all my life!" In its most abstract form, the theory of groups can well be likened to a grin without a cat: the study of the structure of an unspecified relationship between undefined elements.

With the help of the group concept we are able to establish order in what might otherwise be an amorphous jungle of facts. We have already seen how tensors depend on groups for their definition.

Many excellent books on the theory of groups are available. Probably the one by Ledermann[11] will meet most of the requirements of the engineer.

From the study of networks, electrical engineers acquire a smattering of topology. In Chapter 10 we made use of topological methods and introduced the idea of a tree on a network. A more extensive and systematic knowledge of the subject is necessary, however, in order to understand many of the articles on network theory appearing in technical periodicals today.

Books such as those by Veblen,[14] Patterson[15] and Wallace[16] provide very clear and readable introductions to the subject, and for the ambitious reader there is the advanced text on the foundations of algebraic topology by Eilenberg and Steenrod.[18] In his book on topological groups, Pontrjagin[17] includes a lucid, but exceptionally concise, summary of both group theory and topology.

In the field of topology, the theory of groups again offers a helping hand in the task of classifying topological relationships. We have touched on this aspect in connexion with the algebraic diagram of network analysis. A deeper study reveals it to be merely a fragment of a larger and much more intricate mosaic. In particular, the homomorphic mapping of one group on another, as well as the idea of homology sequences, is prerequisite to a thorough understanding of the structure of Kron's methods of network analysis.[20,21]

With a solid knowledge of matrix algebra, and a clear grasp of the methods and scope of tensor analysis, group theory and algebraic topology, an electrical engineer should be adequately equipped to appreciate the pattern of his own subject, and to recognize similar patterns in other fields of science and engineering.

Whether we like it or not, the days of the polyhistor, with his balanced knowledge of men and books, belong to the past, and we are living in an age where, to make his mark, the scientist or engineer must needs strive to know more and more about less and less. This process of differentiation has been carried to such extremes that the specialist is rapidly losing contact with the ramifications of his subject. He speaks a specialized language intelligible only to himself and a few professional cronies, and he does not know whether the work he is doing is not being duplicated in other fields.

Applied with intelligence and care, the geometric methods to which this book serves as an elementary introduction will enable the engineer to catch a glimpse of the panorama of contemporary science, pure and applied, without obscuring his view of his own field.

PROBLEMS

12.1. Calculate $\partial x^i/\partial y^j$ and $\partial y^i/\partial x^j$ for the coordinate transformation

$$x^1 = y^1 \cos y^2$$
$$x^2 = y^1 \sin y^2$$

from Cartesian to polar coordinates. Is the transformation admissible everywhere? Show that $(dy^1, 0)$ and $(0, dy^2)$ are orthogonal.

12.2. Use eqn. (6.7) to derive the metric given in Exercise 1, Section 12.2—

$$g_{ij} = \begin{Bmatrix} 1 & 0 \\ 0 & r^2 \end{Bmatrix}$$

12.3. Verify the statement made in the first paragraph of Section 12.3, namely that a tensor which vanishes in one coordinate system will vanish in all other

coordinate systems of the underlying group, by discussing the transformation of the tensor $a^i_j - a^i_j = 0$.

12.4. Is $a^i_j - b_{ij}$ a tensor? Let $a^i_j - b_{ij}$ vanish in S_x; will the expression also vanish in (say) S_y?

Note in this connexion the manner in which the Cayley–Hamilton theorem was proved: the matrix reduction polynomial was transformed into a coordinate system in which it was readily seen to be equal to a zero matrix. Hence, because of the invertibility of the transformation matrix M, the theorem was shown to be generally valid.

12.5. Discuss Exercise 2, Section 12.2, from a purely matrix point of view (*compare* the Exercises in Sections 6.4 and 7.2).

12.6. Calculate $\partial x^i/\partial y^j$, $\partial y^j/\partial z^k$ and $\partial x^i/\partial z^k$, where the x^i refer to Cartesian, the y^j to cylindrical-polar, and the z^k to spherical-polar, coordinate systems.

12.7. Prove that $a_{,k} = \partial a/\partial x^k$, where a is an invariant, by contracting the divalent, mixed tensor $a^i_{j,k}$ over i and j (*see* eqn. (12.34)).

Bibliography

Algebra—Matrix Algebra

1. Bocher, M., *Introduction to Higher Algebra* (New York, Macmillan, 1907).
2. Birkhoff, G. and MacLane, S., *A Survey of Higher Algebra* (New York, Macmillan, 1953).
3. Perlis, S., *Theory of Matrices* (Cambridge, Mass., Addison-Wesley, 1958).
4. Bodewig, E., *Matrix Calculus* (Amsterdam, North-Holland Publishing Co., 1959).

Tensor Analysis

5. McConnell, A. J., *Applications of Tensor Analysis* (New York, Dover Publications, 1957 (1931)).
6. Eddington, A. S., *The Mathematical Theory of Relativity* (Cambridge University Press, 1954).
7. Sokolnikoff, I. S., *Tensor Analysis* (New York, John Wiley, 1951).
8. Spain, B., *The Tensor Calculus* (Edinburgh, Oliver and Boyd, 1953).
9. Schouten, J. A., *Tensor Analysis for Physicists* (Oxford University Press, 1953).
10. Schouten, J. A., *Ricci Calculus* (Berlin, Springer-Verlag, 1954).

Theory of Groups

11. Ledermann, W., *The Theory of Finite Groups* (Edinburgh, Oliver & Boyd, 1953).
12. Kurosch, A. G., *Gruppentheorie* (Berlin, Akademie Verlag, 1955).

Topology

13. Seifert, H., and Threlfall, W., *Lehrbuch der Topologie* (New York, Chelsea Publishing Co., 1947).
14. Veblen, O., *Analysis Situs* (New York, American Mathematical Society, 1931).
15. Patterson, E. M., *Topology* (Edinburgh, Oliver and Boyd, 1956).
16. Wallace, A. H., *Algebraic Topology* (London, Pergamon Press, 1957).

17. PONTRJAGIN, L., *Topological Groups* (Princeton University Press, 1946).

18. EILENBERG, S., and STEENROD, N., *Foundations of Algebraic Topology* (Princeton University Press, 1952).

Applications of Matrices and Tensors

19. CHARNES, A., *et al.*, *Introduction to Linear Programming* (New York, John Wiley, 1953).

20. KRON, G., *Tensor Analysis of Networks* (New York, John Wiley, 1939).

21. KRON, G., *Tensors for Circuits* (New York, Dover Publications, 1959). (This contains a complete bibliography of Kron's articles and books.)

22. GIBBS, W. J., *Tensors in Electrical Machine Theory* (London, Chapman & Hall, 1952).

23. BEWLEY, L. V., *Tensor Analysis of Electric Circuits and Machines*, (New York, Ronald Press Co., 1961).

24. LE CORBEILLER, P., *Matrix Analysis of Electric Networks* (Cambridge, Mass., Harvard University Press, 1950).

25. KONDO, K., *et al.*, *Memoirs of the Unifying Study of the Basic Problems in Engineering Sciences by Means of Geometry* (Tokyo, Gakujutsu Bunken Fukyu-Kai, Vols. I and II, 1955 and 1958).

Index

ADJUNCT, 29
Affinity, homogeneous, 61
Algebraic diagram, 122
Anti-symmetric matrix, 24
Augmented system matrix, 37

BASE, 11
Basic mesh, 111
Bilinear form, 82

CAYLEY–HAMILTON theorem, 78
Chequerboard array of signs, 22
Christoffel symbols, 151
Class, permutation, 17
Closure of a group, 65
Cocut-set, 111
Cofactor, 20
Cogredience, 60
Combination, linear, 8
Commutative principle, 6, 24
Complementary space, 41
Contraction, 73, 150
Contragredience, 60
Contravariance, 60, 65, 147
Convexity, 100
Coriolis' acceleration, 95
Cotree, 111
Covariance, 60, 65
Cramer's theorem, 34
Cut-set, 111

$\stackrel{D}{=}$: "equal to by definition", 3
Delta, Kronecker, 24, 146
Dependence, linear, 8
Determinant, definition of, 18
Diagonal matrix, 24
Diagonalization, 75
Diakoptics, 135
Differentiation—
  covariant, 153
  intrinsic, 154
Duality of networks, 128
Dummy index, 3

EIGENSPACE, 68
Eigenvalue (latent root), 68
Eigenvector (eigendirection), 68
Equivalence relation (R.S.T. relation), 87

FORM—
  bilinear, 82
  quadratic, 82

GRAPH, 111
Group, 65

IDENTITY matrix (idemfactor), 24, 146
Immittance, 113,
Indefinite form, 84
Independence, linear, 8
Index, 3, 83
  dummy, 3
Intersection, 41
Isomorphism, 66

JACOBIAN, 4, 145

KERNEL-index notation, 3, 7, 148
Kirchhoff's laws, 112
Kronecker delta, 24, 146

LATENT root (eigenvalue), 67
Left inverse, 45
Like matrices, 23
Linear combination, 8, 101
Linear independence, 8
Linear operation, 11
Live index (free index), 148
Lowering an index, 151

MAIN diagonal, 24
Manifold, linear vector, 12
Mapping, linear, 61
Matrix—
  connexion, 114
  diagonal, 24
  form, 82
  inverse, 28, 56
  non-singular, 1
  1-, 2-, 7, 15
  orthogonal connexion, 124
  partitioned, 54
  tie, 139
Metric, 60, 147
Minor, 21
Mixed tensor, 148
Multiplication—
  inner (scalar), 51
  outer (dyadic), 51

NEGATIVE definite form, 84
Node-pair, 110
Normalized vector, 51
Nullity of matrix, 40
Nullmatrix, 23

Nullspace of matrix, 39
Nullvector, 7

ORTHONORMAL matrix, 53

p-DIRECTION (E_p), 12
Positive definite form, 84
Post-factor, 24
Pre-factor, 24
Primitive admittance matrix, 118
Primitive impedance matrix, 116
Product, inner and outer, 51

QUADRATIC form, 82
Quadric surface, 90
Quotient law, 151

RAISING an index, 151
Rank, 12, 16
Reduction polynomial, 68, 72
Right inverse, 28, 46

SCALAR, 6, 51, 150
Scalar product, 51, 60
Simplex method of Dantzig, 103
Skew (anti-symmetric) matrix, 24
Slack variable, 103
Span, 12
Spur (trace), 73, 150
Summation convention, 3, 25, 146
Symmetric matrix, 24, 65, 84
System matrix, 138

TENSOR of inertia, 96
Tie, 136, 139
Trace (*Spur*), 73, 150
Transformation—
  linear, 59
  non-linear, 145
Transpose, 19, 24
Tree, 111

UNION, 41
Unit matrix (idemfactor), 24
Unit vector (normalized vector), 7, 52

VALENCE, 3, 15, 150
Vector, 7
Vector function, linear, 61
Vector manifold, linear, 12

WINDOW, 129

BRAAE MATRIX ALGEBRA for Electrical Engineers PITMAN